Random combinatorial structures

Valentin Féray
Exercises by Jacopo Borga

September 2, 2019

1 Introduction

1.1 General question and motivation

Consider

- a *combinatorial class* $\mathcal{C}$, *i.e.,* a set of combinatorial objects with finitely many objects of each size (permutations, graphs, words, lattice paths);
- for each n, a *probability measure* μ_n on the set C_n of objects of size n; most of the time, this is the uniform distribution;
- a function $f : \mathcal{C} \to \mathbb{Z}_{\geq 0}$ (or $\mathbb{R}$), which we will refer to as *statistics*.

We consider the random variable $X_n = f(\boldsymbol{c}_n)$, where $\boldsymbol{c}_n$ has distribution μ_n.

Question: how does X_n behave asymptotically?

1. asymptotic equivalent for the expectation, the variance, or more generally moments...
2. asymptotics of probability $\mathbb{P}(X_n = k_n)$ for some sequence k_n and/or convergence in distribution after normalization (the first question is in general stronger).
   Auxiliary question: what is the speed of convergence?
3. deviation estimates, e.g. at which speed does $\mathbb{P}(X_n \geq t_n)$ (or $\mathbb{P}(X_n \leq t_n)$) tend to 0 (when it tends to 0)?

Motivation:

- statistics (e.g.: compare "measured" X_n on random permutations with the theoretical value to test the hypothesis "the permutation taken is uniform");

- statistical physics: in statistical physics, we take a microscopic configuration at random among all possible ones and we want to describe behaviour of some macroscopic quantity of the system (which would be the statistics f);
- average analysis of algorithms: here, the combinatorial objects are the input of some algorithm (e.g. a permutation to be sorted) and the statistics is the time (or number of steps) taken by the algorithm (more standard, but not necessarily relevant: worst-case complexity).
- "probabilistic method": prove the existence of combinatorial objects with some given properties by proving that the probability that a random object has these properties is nonzero.

1.2 Number of (short) cycles in random permutations

- Combinatorial class: permutations;
- probability measure: μ_n is the uniform distribution on the set S_n of permutations of size n;
- statistics: total number of cycles X_n, or, for a given $k \geq 1$, number of cycles of length k.

Here are two theorems that we shall prove in the lecture, illustrating convergence in distribution, one towards a discrete limit law, and the other one to a continuous one...

Theorem 1.1 (Goncharov '44)**.** *Fix some integer $k \geq 1$ and define $X_n^{(k)}$ the number of cycles of length k in a random permutation in S_n. Then $X_n^{(k)}$ is asymptotically Poisson with parameter $\frac{1}{k}$, that is: for all i,*

$$\lim_{n\to\infty} \mathbb{P}\big(X_n^{(k)} = i\big) = \frac{e^{-1/k}}{k^i i!}.$$

Theorem 1.2 (Goncharov '44)**.** *The number of cycles X_n of a uniform random permutation in S_n is asymptotically Gaussian with mean and variance* $\log(n)$, i.e.,

$$\lim_{n\to\infty} \mathbb{P}\big(X_n \leq \log(n) + x\sqrt{\log(n)}\big) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-w^2/2} dw.$$

1.3 Longest common substring

- Combinatorial class: pairs (w, w') of *words* of the same length on some finite alphabet $\mathcal{A}$, *i.e.,* the objects of size n are $C_n = (\mathcal{A}^n)^2$;
- probability measures: μ_n is the uniform distribution on C_n (w and w' are both uniform on $\mathcal{A}^n$ and independent);
- statistics: longest common substring $L_n = L(w, w')$ between w and w'.

Theorem 1.3. *Let $\mathcal{A}$ be an alphabet of size k and L_n the longest common substring of two independent uniform random words on the alphabet $\mathcal{A}$. There exists a real number γ_k such that*

$$\mathbb{E}(L_n) \sim \gamma_k n.$$

Moreover,

$$\mathbb{P}(|L_n - \mathbb{E}(L_n)| \geq t_n) \leq 2\exp(-t_n^2/4n).$$

The γ_k are known as the Chvátal-Sankoff constants. Proving their existence is relatively easy (see exercices). Despite efforts in this direction, no formulas for any of the γ_k are known!

1.4 Chromatic number and triangles in random graphs

- Combinatorial class: (simple, undirected, loopless) graphs = pairs (V, E) with $E \subseteq \binom{V}{2}$.
- probability measures: G_n is the Erdős-Rényi random graph $G(n, p)$ (where $p = p(n)$ may depend on n) defined by
  - $V = [n]$
  - for each $e \in \binom{V}{2}$, then $e \in E$ independently with probability p

  Note: if $p = 1/2$, uniform measure on all graphs with vertex set $[n]$.
- statistics: $X_n = \chi(G_n)$ is the chromatic number of G_n.

  Recall that a proper coloring of a graph G is a coloring ot the vertices of G so that no two adjacent vertices share the same color. Then the chromatic number $\chi(G)$ of G is the smallest number of colors needed in a proper coloring of G.

Proposition 1.4 (Shamir–Spencer '87)**.** *For any n, p and $\lambda > 0$, we have*

$$\mathbb{P}\big[\,\big|\chi(G_n) - \mathbb{E}[\chi(G_n)]\big| > \lambda\sqrt{n-1}\,\big] < 2e^{-\lambda^2/2}.$$

For $p = 1/2$, the mean is known to be asymptotically $n/\log(n)$, but this is a harder result than the deviation estimate above.

Other interesting statistics: let T_n be the number of triangles (=triple $\{i, j, k\}$ such that all of $\{i, j\}$, $\{j, k\}$ and $\{i, k\}$ are edges of the graph). A variant is to consider $\mathbf{1}[T_n > 0]$, *i.e.,* we are interested only in the existence of a triangle, not in their precise number.

Theorem 1.5 (Erdős–Rényi '60, Bollobás '81)**.** *Let p_n be a sequence of real number in* $[0, 1]$

1. *if $p_n \ll 1/n$, then $P\big(G(n, p_n)$ contains a triangle$\big) \to 0$;*
2. *if $\lim n\, p_n = c \in (0, +\infty)$, then T_n converge in distribution to a Poisson law of parameter c;*
3. *if $p_n \gg 1/n$, then $P\big(G(n, p_n)$ contains a triangle$\big) \to 1$ and the number of triangles is asymptotically Gaussian.*

We say that there is a threshold at $p_n = 1/n$ for the triangle containment property. Existence and determination of thresholds is a common question in random graph theory.

1.5 Number of prime divisors

An example from number theory... which can be treated by similar methods.

Fix some number n. Take uniformly a random number between 1 and n.
We denote $\nu(x)$ the number of prime divisor (without multiplicities) of x.

Theorem 1.6 (Erdös–Kac '40)**.** *Let n be an integer and x a uniform random number between* 1 *and n. Then $\nu(x)$ is asymptotically Gaussian with mean and variance equivalent to* $\log(\log(n))$.

1.6 Methods

0- Compute explicitly the probability that $\mathbb{P}(X_n = k_n)$, for all k_n, and take limits when n tends to infinity.

Applicable to the number of cycles of fixed length in uniform permutations (using inclusion-exclusion).

1- Compute the characteristic function $\mathbb{E}(e^{itX_n})$ or, setting $u = e^{it}$ the *probability generating function.*

$$\mathbb{E}(u^{X_n}) \sum_{\boldsymbol{c} \in C_n} \mu_n(\boldsymbol{c}) u^{f(\boldsymbol{c})}.$$

Claim: If you can compute this function, you have access to "everything" (moments, convergence in distribution, deviation estimates).

Among the above examples, this is applicable to cycles in permutations, and to number of prime divisors (though in this case, the computation of the generating function is quite involved and we will not use this path in this lecture).

In some cases, this probability generating function can be evaluated through analytic combinatorics, using *bivariate generating functions*.

2- When 1- does not work, one can try to evaluate asymptotically the moments $\mathbb{E}(X_n^r)$. Moments are suited for combinatorial quantities: if X_n counts some substructures in the random object $\boldsymbol{c}_n$, then $X_n = \sum_a \mathbf{1}[a \in \boldsymbol{c}_n]$, and one can expand X_n^r.

Moments can be used to give:

- some deviation estimates, in particular a random variable with a small variance is concentrated around its mean (second moment method);
- some convergence in distribution (moment method);

Among the above examples, this is used for triangles in random graphs, and for the number of prime factors in random integers. It is also applicable to fixed length cycles in permutations.

3- The deviation estimates obtained with moments are usually not good (polynomial decay instead of exponential). We can instead use martingales and Azuma's inequality (even without being able to compute the mean!).

The deviation estimates above for longest common subword and for the chromatic number are **straightforward** *applications of Azuma's inequality.*

1.7 Exercises

Exercise 1.1 (Fekete's lemma). A sequence (u_n) of real numbers is called *superadditive* if and only if:

$$u_{n+m} \geq u_n + u_m, \quad \text{for all} \quad n, m.$$

Show that $\lim_{n\to\infty} \frac{u_n}{n}$ exists (it may be infinite) and is equal to $\sup_{n\geq 1} \frac{u_n}{n}$.

Hint: you may first consider the case where $\frac{u_n}{n}$ has a maximal element $\frac{u_k}{k}$ and consider the Euclidean division of n by k. Then deal with the general case.

Exercise 1.2. Denote $\ell_n^{(k)} = \mathbb{E}[L_n^{(k)}]$ the expectation of the size of the longest common subsequence of two uniform random words of size n on a k-letter alphabet.

1. Using the previous exercise, show that $\ell_n^{(k)}$ is asymptotically equivalent to $\gamma_k n$ for some number γ_k (when n tends to infinity and k is fixed).

2. Show that for any integer k and m,

$$\gamma_{m \cdot k} \leq \gamma_k.$$

3. Recall from the lecture that $\mathbb{P}(|L_n^{(k)} - \mathbb{E}(L_n^{(k)})| \geq t_n) \leq 2\exp(-t_n^2/8n)$, when $t_n \to \infty$. Find $t_n \to \infty$ such that a.s., we have $|L_n^{(k)} - \mathbb{E}(L_n^{(k)})| \leq t_n$ for n large enough.

2 Characteristic functions – the analytic approach

2.1 Basics on characteristic functions

Definition 2.1. Let X a real-valued random variable, its *characteristic function* is defined as

$$\varphi_X(t) = \mathbb{E}(e^{itX}).$$

Notes: - sometimes called *Fourier transform*;
- the expectation above is always well-defined when t is a real number, so φ_X is a function on $\mathbb{R}$ and $|\varphi_X(t)| \leq 1$.

Here are the characteristic functions of some classical distribution

Uniform $X \sim \mathcal{U}([a,b]) \ \to \ \varphi_X(t) = \frac{e^{itb}-e^{ita}}{it(b-a)}$;

Gaussian $X \sim \mathcal{N}(m,\sigma^2) \ \to \ \varphi_X(t) = e^{itm-\sigma^2 t^2/2}$;

Poisson $X \sim \mathsf{Poisson}(\lambda) \ \to e^{\lambda(e^{it}-1)}$;

Geometric (starting at 1) $X \sim \mathsf{Geom}(p) \ \to \frac{pe^{it}}{1-(1-p)e^{it}}$.

Probability generating function (PGF). Assume X takes values on $\mathbb{Z}_{\geq 0}$ (we will always assume this when speaking of PGF). Then we can write

$$\varphi_X(t) = \mathbb{E}(e^{itX}) = \sum_{k=0}^{\infty} \mathbb{P}(X_n = k)e^{itk} = P(e^{it}),$$

where $P(u) := \sum_{k\geq 0} \mathbb{P}(X = k)u^k$ is the PGF of X (well-defined at least for $|u| \leq 1$).

Moments. Moments $\mathbb{E}(X^r)$, if they exist (nothing ensures in general that X^r is in L^1) can be recovered from the characteristic function/PGF by differentiating at $t = 0$ (resp. $u = 1$).

- We have
$$\mathbb{E}(X^r) = \frac{1}{i^r}\frac{d^r}{dt^r}\mathbb{E}(e^{itX})|_{t=0}.$$
- Sometimes, it is more appropriate to take logarithm of φ_X.
$$\mathbb{E}(X) = \frac{1}{i}\frac{d}{dt}\log(\varphi_X(t))|_{t=0}, \quad \mathrm{Var}(X) = \frac{1}{i^2}\frac{d^2}{dt^2}\log(\varphi_X(t))|_{t=0}.$$
Further derivative of $\log(\varphi_X(t))$ are called cumulants of X. They can be expressed in terms of moments, and conversely.
- Differentiating the PGF gives
$$\frac{d^r}{du^r}P(u)|_{u=1} = \mathbb{E}[X(X-1)\dots(X-r+1)].$$
The right-hand side is called factorial moments of X. We can express moments in terms of factorial moments and conversely.

Continuity theorem.

Theorem 2.2. *A sequence $(X_n)_{n\geq 1}$ of random variables tends in distribution to X if and only if, for each $t \in \mathbb{R}$,*

$$\lim \varphi_{X_n}(t) = \varphi_X(t).$$

2.2 First convergence results using characteristic functions

Some examples:

1. Let $F(\sigma)$ denote the set of fixed points of a permutation σ. We consider $X_n = |F(\boldsymbol{\sigma}_n)|$, *i.e.,* the number of fixed points in a uniform random permutation of size n. Its PGF is

$$P_n(u) = \frac{1}{n!} \sum_{\sigma \in S_n} u^{|F(\sigma)|}.$$

Useful trick (related to inclusion-exclusion[1]): set $u = v + 1$ and using $(v+1)^{|F|} = \sum_{E \subseteq F} v^{|E|}$, we have

$$P_n(v+1) = \frac{1}{n!} \sum_{\substack{\sigma \in S_n \\ E \subseteq F(\sigma)}} v^{|E|} \quad = \frac{1}{n!} \sum_{|E| \subseteq [n]} v^{|E|} \#\{\sigma \in S_n : F(\sigma) \supseteq E\}$$

The above cardinality is easily seen to be $(n - |E|)!$ (we permute freely elements in $[n] \setminus E$), so that the summand only depends on $k := |E|$.

$$P_n(v+1) = \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} v^k (n-k)! = \sum_{k=0}^{n} \frac{v^k}{k!}. \tag{1}$$

Clearly, $P_n(v+1) \to e^v$ for all complex numbers v, so, in particular $P_n(u) \to e^{u-1}$ for $u = e^{it}$. Using the continuity theorem we know that X_n tends in distribution to a Poisson law of parameter 1.

Since the limiting distribution is discrete, this means concretely that, for all $i \geq 0$,

$$\mathbb{P}(X_n = i) \to \frac{1}{e\, i!}.$$

2. Let $\kappa(\sigma)$ be the number of cycles of a permutation σ. We consider $X_n = \kappa(\boldsymbol{\sigma}_n)$, *i.e.,* the total number of cycles in a random permutation of size n.

To compute its PGF, we note that $\boldsymbol{\sigma}_n$ can be recursively sampled as follows. Take a uniform random permutation $\boldsymbol{\sigma}_{n-1}$ of size $n-1$ and write it in cycle decomposition. Then

[1] In particular, setting $v = -1$ in Eq. (1) below gives the standard inclusion-exclusion formula for the number of permutations without fixed points (aka derangements).

- for each $i \leq n-1$, with probability $1/n$, we add n right after i in its cycle.
- with probability $1/n$, we add n as a new fixed point.

Claim (straightforward to check): the resulting permutation $\boldsymbol{\sigma}_n$ is uniformly generated.

Note: this recursive way of constructing a uniform permutation of size n is known as the *Chinese Restaurant Process*.

Consequence on PGF:

$$P_n(u) = \left(\tfrac{n-1}{n} + u\,\tfrac{1}{n}\right) P_{n-1}(u).$$

Together with $P_1(u) = u$, we get the following formula for $P_n(u)$

$$P_n(u) = \prod_{j=0}^{n-1} \frac{u+j}{j+1}.$$

For future use, write

$$\log(P_n(u)) = \sum_{j=0}^{n-1} \log(1 + \tfrac{u-1}{j+1}) = \sum_{j=0}^{n-1} \frac{u-1}{j+1} + O\big((u-1)^2\big), \qquad (2)$$

with a constant in the O symbol *not depending on* n. We compute easily

$$\mathbb{E}(X) = \frac{d}{dt} \log(P_n(e^t))|_{t=0} = \sum_{j=1}^{n-1} \frac{1}{j+1} = \log(n) + O(1).$$

$$\mathbb{V}ar(X) = \frac{d^2}{dt^2} \log(P_n(e^t))|_{t=0} = \sum_{j=1}^{n-1} \left(\frac{1}{j+1} - \frac{1}{(j+1)^2}\right) = \log(n) + O(1).$$

We set $X_n^* = \frac{X_n - \log(n)}{\sqrt{\log(n)}}$. Then

$$\log(\varphi_{X_n^*}(t)) = -it\sqrt{\log(n)} + \log(P_n(e^{it/\sqrt{\log(n)}})).$$

Using Eq. (2) and expanding the exponential, we find that, for any fixed real number t,

$$\log(\varphi_{X_n^*}(t)) = -it\sqrt{\log(n)} + \sum_{j=0}^{n-1} \frac{it/\sqrt{\log(n)}}{j+1} + \frac{-t^2/\log(n)}{j+1} + o(1) = -\tfrac{t^2}{2} + o(1).$$

By the continuity theorem, X_n^* converges in distribution to a standard Gaussian random variable.

Quasi-power theorem: We end this section by proving asymptotic normality, whenever the PGF is close to a power.

Theorem 2.3. *Let X_n be a real-valued random variables. Assume that, uniformly on a* complex *neighbourhood of $t = 0$, we have*

$$\mathbb{E}(e^{itX_n}) = \exp(\beta_n U(it) + V(it))(1 + o(1)), \tag{3}$$

where $U(t)$ and $V(t)$ are holomorphic functions with $U(0) = V(0) = 0$ and β_n tends to $+\infty$. Assume finally that $U''(0) \neq 0$. Then we have

$$\mathbb{E}(X_n) = \beta_n U'(0) + O(1);$$
$$\mathrm{Var}(X_n) = \beta_n U''(0) + O(1).$$

Moreover, $X_n^* = \frac{X_n - \beta_n U'(0)}{\sqrt{\beta_n U''(0)}}$ converges in distribution towards a standard Gaussian random variable.

Note: the hypothesis (3) can be equivalently written in terms of the PGF $P_n(u)$: for u in a complex neighbourhood of 1, we have

$$P_n(u) = A(u)\, B(u)^{\beta_n}(1 + o(1)), \tag{4}$$

for some holomorphic function A and B with $A(1) = B(1) = 1$ and

$$B''(1) + B'(1) - B'(1)^2 \neq 0. \quad (\text{"variability condition"})$$

Indeed, we can set $u = e^{it}$ and observe that $A(e^{it})$ and $B(e^{it})$ takes their image in $D(1, 1)$ for t in some neighbourhood of 0 and can therefore be written $A(e^{it}) = \exp(V(it))$ and $B(e^{it}) = \exp(U(it))$, for some analytic functions U and V, namely $U(t') = \log(B(e^{t'}))$ and $V(t') = \log(A(e^{t'}))$. with this change of notation, Eq. (4) and Eq. (3) are equivalent. We can then reexpress the quantity in the theorem as $U'(0) = B'(1)$ and $U''(0) = B''(1) + B'(1) - B'(1)^2$.

Proof. We have

$$\log(\mathbb{E}(e^{itX_n})) = \beta_n\, U(it) + V(it) + o(1).$$

Since we have uniform convergence of analytic functions around $t = 0$, we can derive and keep an $o(1)$ error term (see Proposition A.10). The above estimates for the expectation and the variance follow.

For the convergence in distribution, write (the first equality is obtained as in the computation for the total number of cycles)

$$\begin{aligned}\log(\varphi_{X_n^*}(t)) &= -it\frac{\beta_n U'(0)}{\sqrt{\beta_n U''(0)}} + \log(\mathbb{E}(e^{itX_n}/\sqrt{\beta_n U''(0)}) \\ &= -it\frac{\beta_n U'(0)}{\sqrt{\beta_n U''(0)}} + \beta_n\, U\left(it/\sqrt{\beta_n U''(0)}\right) + V\left(it/\sqrt{\beta_n U''(0)}\right) + o(1).\end{aligned}$$

The $V(\dots)$ term goes to 0 and we can use the expansion $U(s) = U'(0)s + U''(0)s^2 + O(s^3)$, valid for small s. When t is fixed, $s := it/\sqrt{\beta_n U''(0)}$ tends to 0 so that we can use this expansion and find

$$\log(\varphi_{X_n^*}(t)) = -t^2/2 + o(1).$$

We conclude, from the continuity theorem, that X_n^* converges in distribution to a standard Gaussian random variable. □

Applications:

1. Standard central limit theorem: take $(Y_i)_{i\geq 1}$ i.i.d. s.t. $\mathbb{E}[\exp(i\,t\,Y_1)]$ is holomorphic for t in a complex neighbourhood of 0. We have

$$\mathbb{E}(\exp(it(Y_1+\cdots+Y_n))) = \left[\mathbb{E}\,\exp(itY_1)\right]^n.$$

   This is of the form (3) with $\beta_n = n$, $U(it) = \mathbb{E}\,\exp(itY_1)$ and $V(it) = 1$.

   Remark 1: note that the hypothesis put on the distribution of Y_1 is stronger than the usual one in central limit theorem (Y_1 has finite variance). In particular, it implies that Y_1 has finite moments of all orders

   Remark 2: in general, the hypothesis (3) can be thought of as follows: the variable "resembles" a sum of i.i.d. random variables (which explains the asymptotic normality).

2. The asymptotic normality of the number of cycles X_n in random permutations (proved above) can also be obtained using this theorem (see exercise);

3. Examples in the next sections...

2.3 Bivariate generating series

In many cases, $P_n(u)$ is not easy to compute directly but we can compute the *bivariate generating function* (BGF)

$$C(z,u) := \sum_{c\in\mathcal{C}} z^{|c|}u^{f(c)} = \sum_{n,k\geq 0} a_{n,k}z_n u^k.$$

Here c is an object in $\mathcal{C}$, $|c|$ its size and $f(c)$ the statistics of interest. In the second formula, $a_{n,k} = \{c \in C_n : f(c) = k\}$ is the number of objects of size n on which the statistics takes value k. We do not know whether this infinite sum converges for some values of the pair (z,u). The series C can however always be defined as formal power series (i.e. series of the form $\sum_{n,k\geq 0} a_{n,k}z^n u^k$ for some complex coefficients $a_{n,k}$). We will however be interested in cases where it does converge, and z and u can be replaced by (sufficiently small) complex numbers.

From BGF to PGF: The PGF of X_n for each $n \geq 1$ can be recovered from the BGF: if μ_n is the uniform distribution on C_n, then

$$P_n(u) = \frac{[z^n]C(z,u)}{[z^n]C(z,1)}. \tag{5}$$

(if μ_n is not uniform, we would need to put weights in the definition of the BGF; we will not discuss this here).

Cauchy formula for derivatives, and more generally the residue formula, are powerful tools to perform the coefficient extraction in (5).

Product of combinatorial classes: Consider two combinatorial classes $\mathcal{C}$ and $\mathcal{D}$ with some statistics f_C and f_D. Then the product $\mathcal{E} := \mathcal{C} \times \mathcal{D}$ of these classes is defined as the set $\mathcal{C} \times \mathcal{D}$ endowed with the following notion of size: $|(\boldsymbol{c}, \boldsymbol{d})| = |\boldsymbol{c}| + |\boldsymbol{d}|$. (One checks easily that it has finitely many objects of each size.) We also consider the statistics $f_E(c, d) = f_C(c) + f_D(d)$ on $\mathcal{E}$.

Lemma 2.4. *Let $\mathcal{C}$ and $\mathcal{D}$ be combinatorial classes and $\mathcal{E}$ their product, with the above described statistics. Then we have*

$$E(z, u) = C(z, u)\, D(z, u).$$

This always holds as an equality of formal power series. Moreover, if C and D both converge (absolutely) for some values of (z, u), then E also does and the above equality holds.

(Many statements with generating functions should be interpreted this way; I shall not always repeat this.)

Proof. $E(z, u) = \displaystyle\sum_{(c,d)\in\mathcal{E}} z^{|(c,d)|} u^{f_E(c,d)} = \sum_{(c,d)\in\mathcal{E}} (z^{|c|} u^{f_C(c)})\, (z^{|d|} u^{f_D(d)}),$
where the second inequality uses $|(c, d)| = |c| + |d|$
and $f_E(c, d) = f_C(c) + f_D(d)$. □

The sequence operator: Let $\mathcal{C}$ be combinatorial class *without element of size* 0, with some statistics f_C. Then we define

$$\mathrm{Seq}(\mathcal{C}) = \{\varnothing\} \uplus \mathcal{C} \uplus (\mathcal{C} \times \mathcal{C}) \uplus (\mathcal{C} \times \mathcal{C} \times \mathcal{C}) \uplus \dots,$$

where $\varnothing$ is an element of size 0. In other words, an object in $\mathrm{Seq}(\mathcal{C})$ is a finite list (possibly empty) of objects in $\mathcal{C}$. Its size is the sum of the sizes of its components; furthermore, we consider, as statistics f on $\mathrm{Seq}(\mathcal{C})$, the sum of the statistics of its components. Note that $(\mathcal{C})^k$ (k times) has only objects of size at least k (since all objects in $\mathcal{C}$ have size at least 1). This implies for a fixed n, objects of size n in $\mathrm{Seq}(\mathcal{C})$ can only come from the $(\mathcal{C})^k$ for $k \le n$. In particular since each of these $(\mathcal{C})^k$ has finitely many objects of size n, the set $\mathrm{Seq}(\mathcal{C})$ has finitely many of size n and is indeed a combinatorial class.

Lemma 2.5. *Let $\mathcal{C}$ be combinatorial class* without element of size 0 *and set $\mathcal{A} = \mathrm{Seq}(\mathcal{C})$. Then*

$$A(z, u) = \frac{1}{1 - C(z, u)}.$$

$1 - C(z, u)$ is invertible as power series since $C(z, u)$ has no constant term.

Proof. $A(z,u) = 1 + C(z,u) + C(z,u)^2 + C(z,u)^3 + \cdots = \frac{1}{1-C(z,u)}$. □

Example 1: length of the first part in integer compositions An integer composition of n is a list $\boldsymbol{a} = (a_1, \ldots, a_k)$ of positive integers that sum to n. We take $\boldsymbol{a}^{(n)}$ a uniform random composition of n and consider $X_n = (a^{(n)})_1$ its first part. The BGF of *non-empty compositions* with the exponent of u being the first part is

$$C(z,u) = \frac{zu}{1-zu}\,\frac{1}{1-z/(1-z)} = \frac{zu(1-z)}{(1-zu)(1-2z)}.$$

The first fraction corresponds to the first part which contributes $zu + (zu)^2 + (zu)^3 + \ldots$, depending on whether this part is $1, 2, 3, \ldots$ The second fraction corresponds to the other parts: a sequence of elements in $\{1, 2, \ldots\}$, the GF of the later being $z + z^2 + z^3 + \cdots = z/(1-z)$.

For fixed u, the function $C(z,u)$ is meromorphic and has two poles $1/u$ and $1/2$. For u of modulus 1, the pole $1/u$ is bigger than $3/4$ (in modulus) and we can write

$$\oint_{\partial D(0,3/4)} \frac{C(z,u)}{z^{n+1}} dz = 2\pi i \left[\operatorname{Res}(\tfrac{C(z,u)}{z^{n+1}}; 0) + \operatorname{Res}(\tfrac{C(z,u)}{z^{n+1}}; 1/2) \right].$$

The path integral is bounded by $2\pi \sup_{|z|=1} \frac{|C(z,u)|}{(3/4)^{n+1}} = O\big((4/3)^n\big)$. The residue in 0 is $[z^n]C(z,u)$, which is the quantity we're interested in. Since the pole in $1/2$ is simple, we can write

$$\operatorname{Res}(\tfrac{C(z,u)}{z^{n+1}}; 1/2) = \lim_{z \to 1/2} \frac{(z-1/2)C(z,u)}{z^{n+1}}$$

$$= 2^{n+1} \lim_{z \to 1/2} (z-1/2)C(z,u) = \underbrace{(*)}\, 2^{n+1} \operatorname{Res}(C(z,u); 1/2) = -2^n \frac{u/2}{1-u/2}.$$

A factorization, as $(*)$, in the residue of $\frac{C(z,u)}{z^n}$ always occurs when $C(z,u)$ has a simple pole. We get that

$$[z^n]C(z,u) = 2^n \frac{u/2}{1-u/2} + O\big((4/3)^n\big).$$

Going back to the PGF, this gives

$$P_n(u) = \frac{[z^n]C(z,u)}{[z^n]C(z,1)} = \frac{u/2}{1-u/2} + O\big((2/3)^n\big).$$

The fraction $\frac{u/2}{1-u/2}$ is the PGF of the geometric distribution G of parameter $1/2$ (starting at 1), i.e. $\mathbb{P}(G = k) = 1/2^k$. Since the above equation holds for all u of modulus 1, i.e. for all $u = e^{it}$ we have convergence of the characteristic

function of X_n to that of a geometric distribution, and therefore **convergence in distribution of X_n to G.**

Remark. This could be obtained by more elementary means, but this is a good illustration of the combination of BGF, residue theorem and the continuity of the characteristic function.

Example 2: occurrences of the factor ab in random words

Let $\boldsymbol{w}_n$ be a uniform random word of length n on the alphabet $\{a, b, c\}$ (i.e. a uniform random element of $\{a, b, c\}^n$). We are interested in the number $X_n = \#ab(\boldsymbol{w}_n)$ of *consecutive* occurrences (also called factors) of ab in $\boldsymbol{w}_n$.

The combinatorial class of words without statistics is $\mathrm{Seq}(\{a, b, c\})$. Let

$$W(z, u) = \sum_{w \in \{a,b,c\}^*} z^{|w|} u^{\#ab(w)}$$

be the associated BGF. Set $u = v + 1$. Then, denoting $\mathrm{Occ}_{ab}(w)$ the set of occurrences of w,

$$W(z, v+1) = \sum_{\substack{w \in \{a,b,c\}^* \\ E \subset \mathrm{Occ}_{ab}(w)}} z^{|w|} v^{|E|}$$

is the BGF of words with some *marked* occurrences of ab. A word with marked occurrences of ab decomposes as

$$w_0 \underline{ab} w_1 \underline{ab} \dots \underline{ab} w_k,$$

for some k and some words $w_0, \dots, w_k$. In terms of combinatorial class, this is

$$\mathrm{Seq}(\{a, b, c\}) \times \mathrm{Seq}\left(\{\underline{ab}\} \times \mathrm{Seq}(\{a, b, c\})\right).$$

The BGF is

$$W(z, v+1) = \frac{1}{1-3z} \times \frac{1}{1 - vz^2 \cdot \frac{1}{1-3z}},$$

implying

$$W(z, u) = \frac{1}{1 - 3z - (u-1)z^2}.$$

For fixed u in a neighbourhood of 1 but different from 1, this function has two poles

$$z^{\pm}(u) = \frac{3 \pm \sqrt{9 + 4(u-1)}}{-2(u-1)},$$

one converging to $+\infty$ and the other one to $1/3$. Note that here above, we have a square-root of a complex number (closed to 9 when u is close to 1); we interpret that as the principal determination of the square-root, defined on $\mathbb{C} \setminus \mathbb{R}$.

Eventually restricting the neighbourhood of 1 where u lives, one can assume that one is bigger than 1 (in modulus) and the other, that we denote $\rho(u)$, is

smaller. We apply the residue theorem to the function $W(z,u)/z^{n+1}$ on the circle $\partial D(0,1)$.

$$\oint_{\partial D(0,1)} \frac{W(z,u)}{z^{n+1}} dz = 2\pi i \left[\operatorname{Res}(\tfrac{W(z,u)}{z^{n+1}}; 0) + \operatorname{Res}(\tfrac{W(z,u)}{z^{n+1}}; \rho(u)) \right].$$

The path integral is bounded by $2\pi \sup_{|z|=1} |W(z,u)| = O(1)$. The residue in 0 is $[z^n]W(z,u)$, which is the quantity we're interested in. As above, since the pole in $\rho(u)$ is simple, we have

$$\operatorname{Res}(\tfrac{W(z)}{z^{n+1}}; \rho(u)) = \frac{1}{\rho(u)^{n+1}} \operatorname{Res}(W(z); \rho(u)).$$

To sum up we have

$$[z^n]W(z,u) = \frac{-\operatorname{Res}(W(z); \rho(u))}{\rho(u)} \left(\frac{1}{\rho(u)}\right)^n + O(1)$$

The same holds for $u = 1$ (there is only 1 pole $\rho(1) = 1/3$ in this case) and we have

$$P_n(u) = \frac{[z^n]W(z,u)}{[z^n]W(z,1)} = A(u)\, B(u)^n + O(1)$$

with $B(u) = \frac{\rho(1)}{\rho(u)}$. This has the form of the quasi power theorem (Theorem 2.3). Using for example Maple, we get $B'(1) = 1/9$ and $B''(1) = -1/81$, and hence $U'(0) = 1/9$ and $U''(0) = 7/81$. We can apply Theorem 2.3 and get

Proposition 2.6. *The number of factors ab in a random word on the alphabet $\{a, b, c\}$ of length n is asymptotically Gaussian with mean $\frac{1}{9}n$ and variance $\frac{7}{81}n$.*

Discussion:

- *General idea:* If the BGF is meromorphic, extracting the n-th coefficient to compute the PGF can be done with the residue theorem: integrate the function $C(z,u)/z^{n+1}$ on a circle that contains the closest pole(s) to the origin.
- Informally, if the pole does not move with u, then we will have a convergence for fixed u and convergence of the variable without renormalization. If the pole moves with u, then we are in the situation of the quasi-power theorem. (This does not cover all cases; the function could not be meromorphic for some values of u, or the order of the pole could vary, e.g. as for $C(z,u) = \frac{1}{(1-z)(1-uz)}$).

2.4 Labelled classes and exponential BGF

Definition 2.7. A *labelled combinatorial class* is a collection $\mathcal{C} = (C_I)_{I \Subset [n]}$ of sets of combinatorial objects, one for each finite subset of $[n]$. We always assume that C_I is finite and that $|C_I| = |C_J|$ whenever $I = J$.

Its exponential generating function (EGF) is $C(z) = \sum_{n \geq 0} \frac{|C_{[n]}|}{n!} z^n$

Examples: the following formulas define labelled combinatorial class

- $P_I = \{\text{bijections } I \to I\}$, with EGF $P(z) = \sum_n \frac{n!}{n!} z^n = \frac{1}{1-z}$;
- $G_I = \{\text{simple graphs with vertex set } I\}$ with EGF $G(z) = \sum_n \frac{2^{\binom{n}{2}}}{n!} z^n$ (warning: does not converge for $z > 0$, purely formal expression);
- $C_I = \{\text{connected graphs with vertex } I\}$ with EGF $C(z)$ computed later.

The product of labelled combinatorial classes $\mathcal{C} = \mathcal{A} \times \mathcal{B}$ is defined by

$$C_I = \bigsqcup_{\substack{J,K \\ J \sqcup K = I}} A_J \times B_K.$$

In words, an element with label set I in the product is a pair of elements in A and B respectively, with disjoint label sets whose union is K.

Lemma 2.8. *If $\mathcal{C} = \mathcal{A} \times \mathcal{B}$ (as labelled combinatorial class), then their EGF satisfies $C(z) = A(z)\,B(z)$.*

Proof. We note that

$$\frac{1}{n!}|C_{[n]}| = \frac{1}{n!} \sum_{\substack{J,K \\ J \sqcup K = [n]}} |A_J|\,|B_K| = \sum_{\substack{j,k \\ j+k=n}} \frac{1}{j!k!} |A_{[j]}|\,|B_{[k]}|. \tag{6}$$

The second equality is justified as follows: from the definition of labelled combinatorial classes, the summand $|A_J|\,|B_K|$ only depends on the set sizes $j = |J|$ and $k = |K|$, and the number of sets J, K with $J \sqcup K = [n]$ and sizes j and k (with $j + k = n$) is $\binom{n}{j} = \frac{n!}{j!k!}$. Finally (6) translates into $C(z) = A(z)\,B(z)$ (Cauchy product of series). □

Other constructions: let $\mathcal{A}$ be a labelled combinatorial class without elements of size 0

- the class of sequences of elements of $\mathcal{A}$ is

  $$\mathcal{C} = \mathrm{Seq}(\mathcal{A}) := \{\varnothing\} \uplus \mathcal{A} \uplus (\mathcal{A} \times \mathcal{A}) \uplus (\mathcal{A} \times \mathcal{A} \times \mathcal{A}) \uplus \dots,$$

  Its EGF is $C(z) = \frac{1}{1-A(z)}$.

- the class of sets (i.e. unordered sequences) of elements of $\mathcal{A}$ is

  $$\mathcal{C} = \mathrm{Set}(\mathcal{A}) := \{\varnothing\} \uplus \mathcal{A} \uplus \big[(\mathcal{A} \times \mathcal{A})/S_2\big] \uplus \big[(\mathcal{A} \times \mathcal{A} \times \mathcal{A})/S_3\big] \uplus \dots,$$

  The quotient by S_2, S_3 means that we identify sequences in different order (corresponding to the same set). For each $k \geq 1$, the EGF of $(\mathcal{C} \times \cdots \times \mathcal{C})/S_k$ is $A(z)^k/k!$. Indeed, each element of $(\mathcal{C} \times \cdots \times \mathcal{C})/S_k$ corresponds to $k!$ elements in $\mathcal{C} \times \cdots \times \mathcal{C}$ (such elements are k-tuple of objects in $\mathcal{A}$

with disjoint supports, so there cannot be repetitions, ensuring that to a non-ordered elements, corresponds exactly $k!$ ordered versions).

We conclude that the EGF of $\mathrm{Set}(\mathcal{A})$ is

$$C(z) = \sum_{k\geq 0} \frac{A(z)^k}{k!} = \exp(A(z)).$$

Warning: the exponential formula for sets works only in the labelled setting, not in the unlabelled one (because of potential repetitions).

Examples:

- If $\mathcal{G}$ and $\mathcal{C}$ are the class of graphs and connected graphs respectively, we have $\mathcal{G} = \mathrm{Set}(\mathcal{C})$ and hence $G(z) = \exp(C(z))$. This is a purely formal relation (convergent for no z) but nevertheless allows to compute $C(z) = \log(G(z))$.
- Consider the class $\mathcal{C}$ of cyclic permutations. There are $(n-1)!$ cyclic permutations of $[n]$. Its EGF is

$$C(z) = \sum_n \frac{(n-1)!}{n!} z^n = \sum_n \frac{1}{n} z^n = -\log(1-z).$$

Then the class $\mathcal{P}$ of permutations is given by $\mathcal{P} = \mathrm{Set}(\mathcal{C})$ and we find

$$P(z) = \exp(-\log(1-z)) = \frac{1}{1-z},$$

as expected. This way of computing the EGF of permutations is useful when we introduce statistics, see exercises.

Statistics on labelled classes and bivariate EGF A statistics on a labelled combinatorial class is a function $f : C_I \to \mathbb{R}$, for each I, such that, whenever $|I| = |J|$, f is compatible with some bijection $C_I \to C_J$. (examples: number of cycles in permutations, number of connected components/ of triangles in graphs).

Bivariate EGF (for $\mathbb{N}_0$-valued statistics):

$$C(z,u) = \sum_{n,k\geq 0} \frac{a_{n,k}}{n!} z^n u^k,$$

where $a_{n,k} = \#\{c \in C_{[n]} : f(c) = k\}$.

The formulas for EGFs of products, sequences, sets of labelled combinatorial classes are valid for bivariate EGF, as soon as the statistics is additive (i.e. statistics in the product/sum/set class is the sum of the statistics on the original class).

The relation PGF-BGF also holds in the exponential/labelled setting

$$P_n(u) = \frac{[z^n]C(z,u)}{[z^n]C(z,1)}.$$

Application: descents in uniform random permutations A *descent* in a permutation σ of n in a permutation is an integer $i \leq n-1$ such that $\sigma_{i+1} < \sigma_i$. Let $C(z,u)$ the bivariate EGF of permutations, where the exponent of u counts the number of descents.

Step 1: computing $C(z,u)$: As before, we set $u = v+1$. Then $D(z,v) := C(z,v+1)$ is the EGF of permutations with marked descents, where the exponent of v counts the number of marked descents.

Example of a permutation with marked descents: $7\,\underline{8\,5\,3}\,2\,\underline{4\,1}\,\underline{9\,6}$. The underline blocks of marked descents are decreasing segments. The bivariate EGF of these decreasing segments (with statistics length minus 1, which is the number of marked descents in such a segment), including the trivial one of length 1 (non underlined in the example) is $\sum_{n\geq 1} \frac{1}{n!} v^{n-1} z^n = \frac{e^{zv}-1}{v}$.

A permutation with marked descents is an (ordered) sequence of such decreasing segments, so that their bivariate EGF is

$$D(z,v) = \frac{1}{1-\frac{e^{zv}-1}{v}} = \frac{v}{v-e^{vz}+1},$$

from which we get

$$C(z,u) = \frac{u-1}{u-e^{(u-1)z}}$$

Step 2: analysis via residue theorem. For fixed u in a complex neighbourhood of 1, the function $C(z,u)$ is meromorphic with only simple poles at locations

$$\rho_k(u) = \frac{\log(u) + 2k\,\pi\,i}{u-1}, \qquad k \in \mathbb{Z}.$$

Note that $\rho_0(u)$ tends to 1, when u tends to 1, while all other poles tend to infinity. In particular, taking u in a sufficiently small neighbourhood of 1 (in fact $|u-1| < 1/2$ is good here), we can assume that $|\rho_0(u)|$ is smaller than 2 all other poles are bigger than 2.

Residue theorem tells us

$$\frac{1}{2\pi i}\oint_{\partial D(0,2)} \frac{C(z,u)dz}{z^{n+1}} = [z^n]C(z,u) + (\rho_0(u))^{-n-1}\,\mathrm{Res}(C(z,u),\rho_0(u)).$$

A simple computation gives $\mathrm{Res}(C(z,u),\rho_0(u)) = -1/u$. Using the standard estimates, the integral is $O(2^{-n})$ uniformly for u in $D(1,1/2)$. Thus, we get

$$[z^n]C(z,u) = \frac{1}{u}(\rho_0(u))^{-n-1} + O(2^{-n}),$$

Dividing by $[z^n]C(z,1) = 1$ (why?), we get the probability generating function is, uniformly for u in $D(1,1/2)$

$$P_n(u) = \frac{[z^n]C(z,u)}{[z^n]C(z,1)} = \frac{1}{u}(\rho_0(u))^{-n-1} + O(2^{-n}).$$

This is of the form of the quasi-power theorem with $A(u) = (u\rho_0(u))^{-1}$, $B(u) = (\rho_0(u))^{-1}$ and $\beta_n = n$. After computing the derivatives $B'(0) = 1/2$ and $B''(0) = -1/6$ (painful by hand, better use a computer), we conclude that the number of descents in a uniform random permutation is asymptotically Gaussian with mean $n/2 + O(1)$ and variance $n/12 + O(1)$.

2.5 Concentration inequalities

What are concentration inequalities? Take the example of the number of X_n of descents (but the following discussion holds for any asymptotically Gaussian statistics). The above asymptotic Gaussianity result tells us that, for any fixed t

$$\lim_{n\to\infty} \mathbb{P}(X_n \le n/2 + t\sqrt{n/12}) = \tfrac{1}{2\pi} \int_{-\infty}^{t} e^{-u^2/2} du \tag{7}$$

(Here and in what follows, $A \sim B$ means $\lim A/B = 1$.) What about $\mathbb{P}(X_n \le n/3)$? For any $M < 0$ and n sufficiently large (how large depends on M), we have $n/3 \le n/2 + M\sqrt{n/12}$ and thus

$$\mathbb{P}(X_n \le n/3) \le \mathbb{P}(X_n \le n/2 + M\sqrt{n/12}) \to \tfrac{1}{2\pi} \int_{-\infty}^{M} e^{-u^2/2} du,$$

so that

$$\limsup_{n\to\infty} P(X_n \le n/3) \le \tfrac{1}{2\pi} \int_{-\infty}^{M} e^{-u^2/2} du.$$

Since this holds for all $M < 0$, we have

$$\lim_{n\to\infty} P(X_n \le n/3) = 0.$$

This is a first answer, but how fast is this convergence to 0? Upper bounds on this convergence rate are called *concentration inequalities*.

From PGF to concentration inequalities: Chernoff bounds

Lemma 2.9 (Chernoff bounds)**.** *Let X be a real-valued random variable. Then for each a in $\mathbb{R}$ and any positive $u > 1$, we have*

$$\mathbb{P}(X \ge a) \le \frac{\mathbb{E}[u^X]}{u^a}; \qquad \mathbb{P}(X \le a) \le \frac{\mathbb{E}[u^{-X}]}{u^{-a}}.$$

Note: the LHS does not depend on t, so we can choose u as we want to optimize the bound.

Proof. For any nonnegative random variable Y and a nonnegative $s \ge 0$, we have $Y \ge s\mathbf{1}[Y \ge s]$ a.s. which implies

$$\mathbb{E}(Y) \ge s\mathbb{P}[Y \ge s].$$

(This is called Markov's inequality.) Chernoff bounds is simply Markov's inequality applied to $Y = u^X$ (resp. $Y = u^{-X}$), observing that by monotonicity of the exponential function we $\mathbb{P}(X \geq a) = P(u^X \geq u^a)$ (resp. $\mathbb{P}(X \leq a) = P(u^{-X} \geq u^{-a})$). □

Back to the example. For u in $D(1, 1/2)$, we have

$$E[u^{-X_n}] = P_n(1/u) = u(\rho_0(1/u))^{-n-1} + O(2^{-n}).$$

From Chernoff bounds, we get

$$\begin{aligned}\mathbb{P}(X_n \leq n/3) &\leq \frac{u(\rho_0(1/u))^{-n-1} + O(2^{-n})}{u^{-n/3}} \\ &\leq \frac{u}{\rho_0(1/u)} \left(\frac{u^{1/3}}{\rho_0(1/u)}\right)^n (1 + o(1)).\end{aligned}$$

Setting $u = 1.4$ gives $\mathbb{P}(X_n \leq n/3) \leq 2 \times .95^n(1+o(1))$ (with help of a computer). We conclude that the probability that a random permutation has less than $n/3$ descents decays exponentially fast in n (and we could optimize the rate by using other values of u).

2.6 Discussion

- The "analytic residue method" has been presented in this lecture only for meromorphic functions. In fact it work for other functions, e.g. having square root singularity as the Catalan generating functions (though with much bigger error terms).
- From generating functions, we can get other probabilistic estimates, such as lower bounds for concentration inequality (giving large deviation results), estimate of the speed of convergence in the central limit theorem or local limit theorem (describing the probability $\mathbb{P}(X_n = k_n)$ for a given k_n).

2.7 Exercises

Exercise 2.1. The purpose of this exercise is to show that the number C_2^n of cycles of length 2 in a random uniform permutation σ^n of size n is also asymptotically Poisson, with a parameter to be determined.

Denote, for a two element set $I = \{i, j\} \subseteq \{1, \ldots, n\}$,

$$\delta_I(\sigma) = \begin{cases} 1 & \text{if } \sigma(i) = j \text{ and } \sigma(j) = i; \\ 0 & \text{otherwise.} \end{cases}$$

- For distinct 2-element sets $I_1, \ldots, I_k$ compute

$$\mathbb{E}\big(\delta_{I_1}(\sigma^n) \cdots \delta_{I_k}(\sigma^n)\big).$$

Hint: beware that *distinct* 2-element sets are not necessarily *disjoint*.

- How many sets $\{I_1, \dots, I_k\}$ of k pairwise disjoint two elements subsets are there ?
- Compute the probability $P(C_2^n = 0)$.
- For a given integer r, compute $P(C_2^n = r)$ and conclude.
- (Bonus question) Fix a positive integer k. Show that the number C_k^n of cycles of length k in a random uniform permutation of size n is also asymptotically Poisson, with a parameter to be determined.

Exercise 2.2. Let P_n be the PGF of the number of fixed points in a uniform permutations of size n.

- Prove that

$$P_n(u) = P_{n-1}(u) + \frac{u-1}{n}(P_{n-1}(u) - P'_{n-1}(u)).$$

- Deduce, by induction the formula given in the lecture

$$P_n(u) = \sum_{k=0}^{n} \frac{(u-1)^k}{k!}.$$

Exercise 2.3. Let

$$f(z) = e^{z-2} + \frac{3\sin\big(z-2\big)}{\big(z-2\big)^2} \quad \text{and} \quad g(z) = \frac{1}{\big(z-2\big)^2}.$$

1. Find the type of singularity of f and g around $z = 2$.
2. Compute $\mathrm{Res}\big(\frac{f(z)}{z^{n+1}}, 2\big)$ and $\mathrm{Res}\big(\frac{g(z)}{z^{n+1}}, 2\big)$, for all $n \in \mathbb{N}$.

Exercise 2.4. In this exercise we prove the following nice property of power series with non-negative real coefficients (extremely used in combinatorics).

Theorem (Pringsheim's Theorem). *Let $f(z) = \sum f_n z^n$ a power series of radius of convergence $R \in (0, \infty)$ with* **non-negative real coefficients**. *Then f has a singularity in $z_0 = R$.*

1. By contradiction suppose that f is analytic at R. Therefore it is analytic in a disk $D(R, r)$ centered in R of radius r, for some $r > 0$. Compute the series expansion of f around $z_0 = R - h$ for $0 < h < r/3$ using the "derivative formula for coefficients" applied to $f(z) = \sum f_n z^n$. (Why can we do it?)
2. Substitute $z = R+h$ in the expression obtained in the previous step (Why can we do it?). Find a contradiction.

Exercise 2.5. Prove the following lemma that we already saw during the lecture.

Lemma. Let $\mathcal{C}$ *be combinatorial class* without element of size 0 *and set* $\mathcal{A} =$ *Seq*$(\mathcal{C})$*. Then*

$$A(z,u) = \frac{1}{1 - C(z,u)}.$$

Why $1 - C(z,u)$ is invertible as power series?

Exercise 2.6. The goal of this exercise is to prove a central limit theorem for the number of parts in compositions. We recall that a *composition* of an integer n in *kparts* is a way of writing n as the sum of a sequence of (strictly) positive k integers. Two sequences that differ in the order of their terms define different compositions.

Example. The four compositions of 3 are:

- 1+1+1 (3 parts)
- 2+1 (2 parts)
- 1+2 (2 parts)
- 3 (1 part)

1. Find a way to represent every composition of n in k part as a permutation of n dots and $k-1$ bars.
2. We denote with $\mathcal{C}$ the combinatorial class of compositions, with $\{\bullet\}$ the combinatorial class containing a single dot and with $\{\bullet, |\bullet\}$ the combinatorial class containing a single dot and a bar followed by a dot. Prove that the following relation holds

$$\mathcal{C} = \{\bullet\} \times \mathrm{Seq}\big(\{\bullet, |\bullet\}\big). \tag{8}$$

3. Deduce from the previous equation that the BGF for the number of parts in a composition is

$$C(z,u) = \frac{uz}{1 - z(1+u)}.$$

Hint: What is the BGF for $\{\bullet, |\bullet\}$? Note that in Equation (8) we are "undercounting " by 1 the number of parts.

4. Deduce an expression for the PGF and conclude using the quasi-power theorem.

Hint: Recall that $\frac{1}{1-x} = \sum_{n\geq 0} x^n$.

The following exercise sums up some of the most important properties of the *Gamma function*. You can skip it and assuming all the results in order to solve Exercise 2.8.

Exercise 2.7. For z with $\Re(z) > 0$, we define

$$\Gamma(z) = \int_0^\infty x^{z-1}\, e^{-x} dx.$$

(Recall that $x^w := \exp\big[w \log(x)\big]$ is well defined for complex exponents w, when x is a positive real number.)

1. Justify that the function is well-defined, *i.e.* that the integral is convergent whenever $\Re(z) > 0$.

2. Show that Γ is holomorphic on $\{z \in \mathbb{C} : \Re(z) > 0\}$.

   Hint: use Morera's criterium (recalled below), and Fubini's theorem to exchange the integral over a triangle and the one from 0 to $+\infty$ in the definition of Γ.

   (Morera's criterion) Let $f : U \to \mathbb{C}$ continuous, U open. Assume that, for all triangles $[A; B; C; A]$ that are completely included in U,

   $$\int_{[A;B;C;A]} f(z) dz = 0.$$

   Then f is holomorphic.

3. Prove that for each $z > 0$, one has $\Gamma(z+1) = z\Gamma(z)$. Conclude that $\Gamma(n) = (n-1)!$ when n is a positive integer.

4. Show that Γ admits an analytic extension to $\mathbb{C} \setminus \{0, -1, -2, \dots\}$. (Hint: use the previous question to first extend Γ to $\{z \in \mathbb{C} : \Re(z) > -1\}$, and then make an inductive proof.)

Exercise 2.8. In this exercise we re-derive the asymptotic normality of the number of cycles in a uniform permutation using the quasi-power theorem.

1. Starting from the formula (that we saw during the lecture) for the PGF for the number of cycles in a uniform permutation, show that

   $$P_n(u) = \frac{\Gamma(u+n)}{\Gamma(u)\Gamma(n+1)}.$$

2. Conclude using the Stirling approximation $\Gamma(z+1) \sim \sqrt{2\pi z}\big(\frac{z}{e}\big)^z$ and the quasi-power theorem.

Exercise 2.9. Compute the bivariate *exponential* generating functions for:

1. the number of cycles in permutations,

2. the total length of cycles in permutations.

Deduce the corresponding *probability* generating functions.

Exercise 2.10. The goal of this exercise is to rederive the Poisson asymptotic behavior for the number of cycles of fixed length m (see Exercise 1 in sheet 2).

1. For $m = 1$, show that the bivariate *exponential* generating function $C(z, u)$ for the number of cycles of length one (*i.e.,* fixed points) in permutations is
$$C(z, u) = \frac{\exp(z(u-1))}{1-z}.$$

2. What is the singularity of $C(z, u)$? Using the residue theorem deduce the asymptotic behavior of $[z^n]C(z, u)$.

3. Deduce the asymptotic behavior for the corresponding PGF and conclude using the continuity theorem.

4. Note that the previous result can be also obtained writing an explicit expression for the coefficients $[z^n\, u^k]C(z, u)$. What are this coefficients? How can you deduce the Poisson asymptotic behavior?

5. Deduce the result for every $m \geq 1$. Start by showing that the bivariate *exponential* generating function $C_m(z, u)$ for the number of cycles of length m is
$$C_m(z, u) = \frac{\exp((u-1)\frac{z^m}{m})}{1-z}.$$
Then use similar arguments to the ones used in points 2 and 3.

Exercise 2.11. The goal of this exercise is to derive the asymptotic normality for the number of parts in set-compositions.

We recall that a set composition of size n is an ordered sequence of disjoint and non-empty sets (called parts) whose union is $\{1, 2, \ldots, n\}$.

Example. $(\{1, 3\}, \{7, 5\}, \{4, 6, 8\}, \{2\})$ is a set partition of size 8 in 4 parts.

1. Show that the bivariate *exponential* generating function $C(z, u)$ for the number of parts in set-compositions is
$$C(z, u) = \frac{1}{1 - u(\exp(z) - 1)}.$$

2. What are the singularities of $C(z, u)$? Is $C(z, u)$ meromorphic?

3. Using the residue theorem deduce the asymptotic behavior of $[z^n]C(z, u)$.

4. Deduce the asymptotic normality for the corresponding PGF and conclude using the quasi-power theorem.

Exercise 2.12. Define the unit n-hypercube to be the set of points $[0,1]^n \subset \mathbb{R}^n$. For example, the unit 0-hypercube is a point, and the unit 3-hypercube is the unit cube. Define a k-face of the unit n-hypercube to be a copy of the k-hypercube on the boundary of the n-hypercube. More formally, a k-face of the unit n-hypercube is a set of the form $\prod_{1\leq i\leq n} S_i$, where S_i is either$\{0\}, \{1\}$ or $[0,1]$ for each i between 1 and n and there are exactly k indices i such that $S_i = [0,1]$.

1. What are the number of k-faces in the unit n-hypercube? Derive the corresponding ordinary BGF.
2. Use the ordinary BGF to derive the expected value of the dimension of a random face of the unit n-hypercube.

Exercise 2.13. Let $(Y_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. real-valued random variables. Set $X_n = \sum_{i=1}^n Y_i$. Show that for each a in $\mathbb{R}$ and any positive $u > 1$, we have

$$\mathbb{P}(X_n \geq a) \leq \frac{\mathbb{E}[u^{X_1}]^n}{u^a} \quad \text{and} \quad \mathbb{P}(X_n \leq a) \leq \frac{\mathbb{E}[u^{-X_1}]^n}{u^{-a}}.$$

Exercise 2.14. Let $(Y_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. random variables such that

$$\mathbb{P}(Y_i = 1) = \mathbb{P}(Y_i = -1) = 1/2.$$

Set $X_n = \sum_{i=1}^n Y_i$. Show, using the previous exercise, that for each a in $\mathbb{R}$,

$$\mathbb{P}(X_n \geq a) = P(X_n \leq -a) \leq e^{\frac{-a^2}{2n}}. \tag{9}$$

(Hint: Note that $\frac{e^x+e^{-x}}{2} \leq e^{\frac{x^2}{2}}$.)

Deduce form (9) that if $(Z_i)_{i\in\mathbb{N}}$ is a sequence of i.i.d. random variables such that

$$\mathbb{P}(Z_i = 1) = \mathbb{P}(Z_i = 0) = 1/2,$$

and $S_n = \sum_{i=1}^n Z_i$ then that for each a in $\mathbb{R}$,

$$\mathbb{P}(S_n - \frac{n}{2} \geq a) \leq e^{\frac{-2a^2}{n}}.$$

Exercise 2.15. Let T_n be a *tournament* on n vertices, that is is a directed graph (digraph) obtained by assigning a direction for each edge in an undirected complete graph on n vertices. Given an ordering $\sigma : [n] \to [n]$, we say that i, j form an *upset* if $i \to j$ but $\sigma(i) > \sigma(j)$.

You can think at the following situation: imagine a tennis tournament among n players where every player play a match against all the other players. If the player i beats j we add the directed edge $i \to j$. At the end of the tournament we obtain a complete directed graph, *i.e.* a tournament. Moreover the ATP ranking gives an ordering of the players. The event that the player i won against j but ranked below j in the ATP ranking corresponds to an upset.

1. Show that for every possible tournament on n vertices, there exist an ordering σ such that the number of upsets is at most $\frac{1}{2}\binom{n}{2}$.

2. Let now T_n be a uniform tournament on n vertices and fix an ordering $\sigma : [n] \to [n]$. Set $U_n(\sigma)$ be the number of upsets corresponding to the tournament T_n and the ordering σ. Using the previous exercise show that

$$\mathbb{P}\left(U_n(\sigma) - \frac{1}{2}\binom{n}{2} \leq -a\right) \leq e^{-\frac{4a^2}{n^2}}.$$

3. Setting $a = n^{\frac{3}{2}}\sqrt{\log(n)}$ deduce that

$$\mathbb{P}\left(\exists \sigma \text{ s.t. } U_n(\sigma) \leq \frac{1}{2}\binom{n}{2} - n^{\frac{3}{2}}\sqrt{\log(n)}\right) \leq n! n^{-4n}. \tag{10}$$

4. Conclude that there exist a tournament T on n vertices such that for every possible ordering $\sigma : [n] \to [n]$, the number of upsets in T is at least $\frac{1}{2}\binom{n}{2} - n^{\frac{3}{2}}\sqrt{\log(n)}$.

Remark. Note that we proved a deterministic result using a probabilistic proof.

Remark. The exponential bounds obtained in Exercise 2.14 are very useful for the union bound in Equation (10).

3 First and second moment method

In cases where we cannot compute the characteristic functions, one can obtain information by computing moments.

3.1 First moment method

The *first moment method* is used to prove that a random structure does not contain a given substructure with high probability. We do this by computing the *expectation (or first moment)* of the number of such substructures and by using on the following simple lemma:

Lemma 3.1. *Let X be a random variable with non-negative integer values (for example X is counting something). Then*

$$P(X = 0) > 1 - \mathbb{E}(X).$$

Proof. $\mathbb{E}(X) = \sum_k kP(X = k) \geq \sum_{k\geq 1} P(X = k) = 1 - P(X = 0)$. □

Consequently, if X_n is a sequence of r.v. with $\lim \mathbb{E}(X_n) = 0$, then $P(X_n = 0)$ tends to 1.

Number of triangles in $G(n, p_n)$.

Reminder from Section 1.4: $G(n, p)$ is the graph with vertex-set $[n] := \{1, \dots, n\}$ and, for each pair $\{i, j\} \subset [n]$, an edge between i and j independently with probability p.

Let p_n be a $[0, 1]$-valued sequence. We are interested in the number T_n of triangles in $G(n, p_n)$. By definition, a triple $I = \{i, j, k\} \subset [n]$ (i.e. $I \in \binom{[n]}{3}$) is a triangle in a graph G if $\{i, j\}$, $\{i, k\}$ and $\{j, k\}$ all belong to the edge-set E_G. We have

$$\begin{aligned} T_n &= \sum_{I\in\binom{[n]}{3}} \mathbf{1}\big[I \text{ is a triangle in } G(n, p_n)\big] \\ &= \sum_{I=\{i,j,k\}\in\binom{[n]}{3}} \mathbf{1}[\{i, j\} \in E_G]\, \mathbf{1}[\{i, k\} \in E_G]\, \mathbf{1}[\{j, k\} \in E_G] \end{aligned}$$

We take expectations: by definition, the random variables $\mathbf{1}[e \in E_G]$ all have expectation p_n and are independent. Therefore each summand above have expectation p_n^3. Since there are $\binom{n}{3}$ summands, we get

$$\mathbb{E}(T_n) = \binom{n}{3} p_n^3.$$

When $p_n = o(n^{-1})$, we have that $\mathbb{E}(T_n)$ tends to 0, and therefore that $P(T_n = 0)$ tends to 1. This proves the first item in Theorem 1.5, which we restate here for convenience:

Proposition 3.2. *If $p_n \ll 1/n$, then $P\big(G(n, p_n)$ contains a triangle$\big) \to 0$;*

Cliques, independent sets and Ramsey numbers A clique in a graph G is a subset I of the vertices, such that any two vertices of I are connected by an edge. An independent set is a subset I of the vertices, so that no pair of vertices in I are linked by an edge. We consider the random variable X_n^k, which is the total number of cliques and independent sets of size k in a uniform random graph with vertex-set $[n]$ (i.e. in $G(n, 1/2)$). A similar computation as above gives

$$\mathbb{E}(X_n^k) = \binom{n}{k} \left(\frac{1}{2}\right)^{\binom{k}{2}}.$$

Theorem 3.3 (Erdős, 1947)**.** *Let $n = \left\lfloor \frac{k}{e\sqrt{2}} 2^{k/2} \right\rfloor$. Then for k large enough, there exists a graph with n vertices and no cliques or independent sets of size k.*

Proof. As k tends to infinity, with n given as above, we have ($A \sim B$ means $\lim A/B = 1$):

$$\mathbb{E}(X_n^k) \sim \frac{n^k}{k!} \left(\frac{1}{2}\right)^{\binom{k}{2}} \sim \frac{1}{2\pi k} \to 0.$$

The first step uses that $\binom{n}{k} \sim \frac{n^k}{k!}$, which is valid when n and k tend simultaneously to infinity, *as long as* $k = o(\sqrt{n})$ (exercise!). The second step uses the formula for n and Stirling formula.

Using the above lemma, we conclude that $\mathbb{P}(X_n^k) = 0$ tends to 1 as k tends to infinity and $n = \left\lfloor \frac{k}{e\sqrt{2}} 2^{k/2} \right\rfloor$. In particular, there exists a graph G with $X_n^k(G) = 0$, i.e. with no cliques or independent sets of size k. □

Historical/general comments. This theorem is interesting because of an earlier result of Ramsey (1930): for every k, there exists $R(k, k)$ such that every graph with at least $R(k, k)$ contains either a clique or an independent set of size k. W.l.o.g., we can assume $R(k, k)$ minimal with this property. The question is then: what is the value of $R(k, k)$?

Erdős theorem shows that $R(k, k) > \frac{k}{e\sqrt{2}} 2^{k/2}$, for large k. The best current known asymptotic lower bound is only twice Erdős' lower bound (even if this is a famous and widely studied question).

It might seem non-natural to consider random graphs to find a graph of size n without cliques or independent sets of size k (the question is not probabilistic in nature!). However, a deterministic construction of such a graph is not known for $n = \left\lfloor \frac{k}{e\sqrt{2}} 2^{k/2} \right\rfloor$ (in fact, not for any n growing exponentially fast with respect to k). So that considering random graphs is in fact a wonderful idea (Erdős was the first to study random graphs, precisely in this context). Considering random objects can be used to show the existence of various combinatorial objects with given properties (without constructing them explicitly): this is known as the *probabilistic method*.

Increasing subsequences in random permutations: Consider a uniform random permutation of size n and L_n the length of its largest increasing subsequence.

Proposition 3.4.

$$P(L_n \geq 3\sqrt{n}) \to 0.$$

Proof. Let X_n be the number of increasing subsequences of size $3\sqrt{n}$ in a uniform random permutation (to simplify notation, we suppose that $3\sqrt{n}$ is an integer; for the general case, replace $3\sqrt{n}$ by $\lceil 3\sqrt{n} \rceil$):

$$X_n = \sum_{I \in \binom{[n]}{3\sqrt{n}}} \mathbf{1}[\sigma/I \text{ is increasing }].$$

Fix $I \in \binom{[n]}{3\sqrt{n}}$. Then

$$\mathbb{P}[\sigma/I \text{ is increasing }] = \frac{1}{n!} \#\{\sigma \in S_n : \sigma/I \text{ is increasing }]\}.$$

The numerator can be computed as follows: choose the *set* of values of σ/I ($\binom{n}{3\sqrt{n}}$ choices; since σ/I is increasing, this determines entirely σ/I), and choose how to complete the permutations (still $n-3\sqrt{n}$ elements to send in all possible ways to $n-3\sqrt{n}$ values, thus $(n-3\sqrt{n})!$ choices). To sum up,

$$\mathbb{P}[\sigma/I \text{ is increasing }] = \frac{\binom{n}{3\sqrt{n}}(n-3\sqrt{n})!}{n!} = \frac{1}{(3\sqrt{n})!}$$

and

$$\mathbb{E}(X_n) = \binom{n}{3\sqrt{n}} \frac{1}{(3\sqrt{n})!} = \frac{n!}{(n-3\sqrt{n})!(3\sqrt{n})!^2}.$$

Using Stirling formula, we have

$$\mathbb{E}(X_n) \sim \frac{n^n e^{3\sqrt{n}}}{(n-3\sqrt{n})^{n-3\sqrt{n}}(9n)^{3\sqrt{n}}(6\pi\sqrt{n})}.$$

Noting that

$$(n-3\sqrt{n})^{n-3\sqrt{n}} = n^{n-3\sqrt{n}}(1-3n^{-1/2})^{n-3\sqrt{n}} = n^{n-3\sqrt{n}}\, e^{-3\sqrt{n}}\, O(1),$$

we have

$$\mathbb{E}(X_n) \sim \frac{(e^2)^{3\sqrt{n}}}{9^{3\sqrt{n}}(6\pi\sqrt{n})} \to 0.$$

This implies that $X_n = 0$ with probability tending to 1, i.e. that, with probability tending to 1, a uniform random permutation has no increasing subsequence of length $3\sqrt{n}$ (i.e. $L_n \leq 3\sqrt{n}$). □

Historical/general comments. The study of largest increasing subsequences in (random) permutations leads to beautiful mathematics, using surprisingly many different methods, see
D. Romik, *The Surprising Mathematics of Longest Increasing Subsequences*, Cambridge University Press, 2015.

Remarks:

- in all examples, we write the r.v. of interest as a sum and use the linearity of expectation (expectation is linear, regardless of independence, which is good!).
- Using only first moments, one cannot prove *the existence* of some object with high probability ;

  $$\mathbb{E}(X_n) \to \infty \text{ does NOT imply } P(X_n = 0) \to 0.$$

  Indeed, consider the sequence X_n, such that $X_n = n^2$ with probability $1/n$ and 0 otherwise; it satisfies $\mathbb{E}(X_n) \to \infty$ and $P(X_n = 0) \to 1$.

3.2 Second moment method

The second moment method allows to prove that a random object contains a given substructure with high probability (and more precisely, that the number of copies of that substructure is concentrated around its expectation.) It relies on computing the variance of this number of substructures and uses the following simple lemma:

Lemma 3.5 (Chebyshev's inequality)**.**

$$\mathbb{P}\Big(|X - \mathbb{E}(X)| \geq \lambda\sqrt{\mathrm{Var}(X)}\Big) \leq 1/\lambda^2.$$

In particular,

$$\mathbb{P}(X = 0) \leq \left(\frac{\sqrt{\mathbb{E}(X)}}{\mathrm{Var}(X)}\right)^2.$$

Proof. Again this uses Markov's inequality, which we recall for convenience: if Y is a non-negative r.v., then $P(Y \geq a) \leq \mathbb{E}(Y)/a$.

We apply this to $Y = (X - \mathbb{E}(X))^2$ and $a = \lambda^2 \mathrm{Var}(X) = \lambda^2\mathbb{E}(Y)$. The result follows immediately. □

Consequently, if X_n is a sequence of r.v. s.t. $\mathrm{Var}(X) = o\big(\mathbb{E}(X)^2\big)$, then $\mathbb{P}(X_n = 0) \to 1$ and more precisely $\frac{X_n}{\mathbb{E}(X_n)} \to 1$ in probability (use Chebyshev's inequality with $\lambda = \varepsilon\frac{\mathbb{E}(X)}{\sqrt{\mathrm{Var}(X)}}$).

Application to triangles in Erdős-Rényi random graph.

As above, let T_n be the number of triangles in $G(n, p_n)$. We recall that $T_n = \sum_{I \in \binom{[n]}{3}} \Delta_I$, where Δ_I is the indicator function of the event "the triangle I is in the graph".

We have computed the expectation $\mathbb{E}(T_n) = \binom{n}{3}p_n^3$. Let us consider the variance.

$$\mathrm{Var}(T_n) = \mathrm{Var}\left(\sum_{I \in \binom{[n]}{3}} \Delta_I\right) = \sum_{I,J \in \binom{[n]}{3}} \mathrm{Cov}\left(\Delta_I, \Delta_J\right).$$

- If I have J have at most 1 element in common, then the corresponding triangles are edge-disjoint and the r.v. Δ_I and Δ_J are independent; the covariance $\mathrm{Cov}\left(\Delta_I, \Delta_J\right)$ is therefore 0.

- if I have J have exactly one edge in common, then

$$\mathrm{Cov}\left(\Delta_I, \Delta_J\right) = \mathbb{E}(\Delta_I\, \Delta_J) - \mathbb{E}(\Delta_I)\mathbb{E}(\Delta_J) = p_n^5 - p_n^6 = p_n^5(1-p_n).$$

  There are $\binom{n}{3}\, 3\,(n-3)$ such terms (choose I, choose which edge it shares with J, choose the last vertex in J).

- If $I = J$ then

$$\mathrm{Cov}\left(\Delta_I, \Delta_J\right) = \mathrm{Var}(\Delta_I) = \mathbb{E}(\Delta_I) - \mathbb{E}(\Delta_I)^2 = p_n^3(1-p_n^3).$$

  There are $\binom{n}{3}$ such terms in the above sum.

Finally we have

$$\mathrm{Var}(T_n) = \binom{n}{3}\, 3\,(n-3) p_n^5(1-p_n) + \binom{n}{3} p_n^3(1-p_n^3).$$

From now on, we assume $n\,p_n \to +\infty$ as $n \to \infty$. We have

$$\frac{\mathrm{Var}(T_n)}{E(X)^2} = O\left(\frac{1}{n^2\, p_n} + \frac{1}{n^3 p_n^3}\right) = o(1).$$

We conclude that, when $n\,p_n \to +\infty$, the random graph $G(n, p_n)$ contains a triangle with probability tending to 1 (this was the first part of item 3. of Theorem 1.5; the Gaussian fluctuations are to be proved later in the lecture).

Descents in uniform random permutations

Let X_n be the number of descents in a uniform random permutations. (We forget the results proved with generating functions; this is an alternate approach, shown on the same example to compare the methods.) We write $X_n = \sum_{i=1}^{n-1} D_i$, where $D_i(\sigma) = 1$ if σ has a descent in position i and 0 otherwise.

To compute its moments, we will use the following construction. Take n i.i.d. uniform random variables $U_1, \ldots, U_n$ in $[0,1]$ and let σ be the (random) permutation such $\sigma_i = j$ is U_i is the j-th smallest value in $\{U_1, \ldots, U_n\}$.

Claim: σ is uniformly distributed.

Proof of the claim. Since transpositions generate the symmetric group, it is enough to prove that σ and $\tau = \sigma \circ (i, i')$ are identically distributed for any transposition (i, i') in S_n. But if σ is associated with $U_1, \ldots, U_n$ by the above procedure, then τ is associated with the list $V_1, \ldots, V_n$ obtained from $U_1, \ldots, U_n$ by swapping U_i and $U_{i'}$. But, for fixed (i, i'), the vectors $(U_1, \ldots, U_n)$ and $(V_1, \ldots, V_n)$ are identically disctributed, implying that σ and τ are identically distributed, as wanted.

Back to descents. Note that, with the above construction, $D_i(\sigma) = 1$ if and only if $U_i > U_{i+1}$. Using this, we can compute the first moments of X_n. First,

$$\mathbb{E}(X_n) = \sum_{i=1}^{n-1} \mathbb{E}(D_i) = \frac{n-1}{2}.$$

Indeed, $\mathbb{E}(D_i) = \mathbb{P}(D_i = 1) = \mathbb{P}(U_i > U_{i+1}) = 1/2$. Furthermore

$$\mathrm{Var}(X_n) = \sum_{1 \le i,j \le n-1} \mathrm{Cov}(D_i, D_j).$$

- If $|i-j| > 1$, then D_i and D_j are independent. With the above representation of σ, the r.v. D_i depends on $\{X_i, X_{i+1}\}$, while D_j depends on the disjoint set $\{X_j, X_{j+1}\}$.
- If $j = i+1$ or $i = j+1$, then

  $$\mathrm{Cov}(D_i, D_j) = \mathbb{E}(D_i\, D_j) - \mathbb{E}(D_i)^2 = \frac{1}{6} - \frac{1}{4} = -\frac{1}{12}.$$

  For $\mathbb{E}(D_i\, D_j)$, we observe that (assuming w.l.o.g. $j = i+1$) $D_i\, D_j = 1$ if $\sigma(i) > \sigma(i+1) > \sigma(i+2)$ (or equivalently, $U_i > U_{i+1} > U_{i+2}$) and 0 otherwise; this implies $\mathbb{E}(D_i\, D_j) = 1/6$ (all 6 possible relative orders of $\sigma(i), \sigma(i+1), \sigma(i+2)$, or equivalently U_i, U_{i+1}, U_{i+2} are equally likely), as claimed above. There are $2(n-2)$ such terms.
- It $i = j$, then $\mathrm{Cov}(D_i, D_j) = \mathrm{Var}(D_i) = \frac{1}{4}$. There are n such terms.

We conclude that

$$\mathrm{Var}(X_n) = -\tfrac{1}{12} \cdot 2(n-2) + \tfrac{1}{4}\, n = \tfrac{1}{12} n + \tfrac{1}{6}.$$

Clearly, $\mathrm{Var}(X_n) = o\big(\mathbb{E}(X_n)^2\big)$ and we conclude that $\frac{X_n}{\mathbb{E}(X_n)}$ tends to 1 in probability (the fact that $\mathbb{P}(X_n = 0)$ tends to 0 is trivial in this example). More precisely, we have the concentration inequality

$$\mathbb{P}\big[|X_n - \mathbb{E}(X_n)| \ge \varepsilon \mathbb{E}(X_n)\big] \le \varepsilon^{-2} \left(\frac{\mathrm{Var}(X)}{\mathbb{E}(X)^2} \right) \sim \frac{\varepsilon^{-2}}{3n}.$$

This concentration inequality is less good that the one obtained through characteristic function (we saw that such quantities decay exponentially fast in n). However computing the variance is easier than computing the full probability generating functions (e.g. for triangles in random graphs, we do not know how to compute the PGF), so that Chebyshev's inequality is a good tool to get some first nonsharp concentration inequalities.

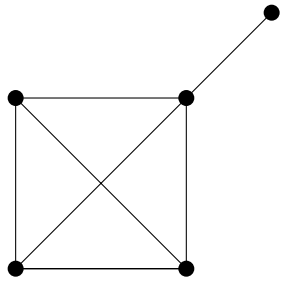
3.3 Exercises

Exercise 3.1. Show that with probability tending to one, a uniform random permutation does not contain three consecutive entries with three consecutive increasing values (i.e. there is no i s.t. $\sigma_{i-1} + 1 = \sigma_{i+1} - 1 = \sigma_i$).

Exercise 3.2. Using the Stirling's formula, prove the claim from the lecture: If n and k tend simultaneously to infinity with $k = o(\sqrt{n})$, then we have

$$\binom{n}{k} \sim \frac{n^k}{k!}.$$

Exercise 3.3. We denote with K_4 a 4 clique. The graph H is obtained from K_4 by adding an extra vertex and edge linking this new vertex to some vertex in K_4 as shown below:



Let X_{K_4} and X_H be the number of copies of K_4 and H in the Erdős-Rényi graph $G(n, p_n)$, respectively.

1. Show that for $p_n \ll n^{-4/6}$ then $\mathbb{E}(X_{K_4})$ tends to zero. What can you conclude?
2. Show that for $p_n \gg n^{-5/7}$ then $\mathbb{E}(X_H)$ tends to infinity.
3. Find a range of values of p_n for which $\mathbb{E}(X_H)$ tends to infinity, but the probability $\mathbb{P}(X_H > 0)$ tends to 0?

Exercise 3.4 (Finiteness of Ramsey's numbers). Prove that, for every $t, s > 0$, there exists $R(t, s)$ such that every graph with at least $R(t, s)$ vertices contains either a clique of size t or an independent set of size s. We recall that the minimal $R(t, s)$ with this property is called Ramsey number.

Hint: prove that $R(t, s) \leq R(t-1, s) + R(t, s-1)$. For that it's convenient to consider a graph G with $R(t-1, s) + R(t, s-1)$ vertices, an arbitrary vertex v in such a graph and the induced graph on $N(v)$ (neighborhood of v) and $V \setminus \{N(v) \cup \{v\}\}$. Then...

Exercise 3.5. We say that a set $A = \{a_1, \dots, a_k\} \subset [n]$ is *sum-free* if for all $S \subset [k]$ the sums $\sum_{i \in S} a_i$ are distinct. Let $K(n)$ denotes the maximal cardinality of a sum-free set contained in $[n]$.

1. Show that for every $n \geq 1$,
$$K(n) \geq \log_2(n).$$
2. Show that for every $n \geq 1$, $K(n)$ satisfies the inequality
$$K(n) \cdot n \geq 2^{K(n)}.$$
3. Deduce that
$$K(n) \leq \log_2(n) + \log_2(\log_2(n)) + cost.$$
4. Fix a sum-free set $A = \{a_1, \dots, a_k\} \subset [n]$ of cardinality k. Consider now the following random variable
$$X = \sum_{i=1}^{k} \varepsilon_i a_i,$$
where ε_i are i.i.d. Bernoulli random variables of parameter $1/2$. Show that
$$\mathbb{P}(|X - \mathbb{E}[X]| \geq n\sqrt{k}) \leq \frac{1}{4}.$$
5. Deduce that
$$\frac{3}{4} 2^k \leq 2n\sqrt{k}.$$
6. Conclude that
$$K(n) \leq \log_2(n) + \frac{1}{2} \log_2(\log_2(n)) + cost.$$

Exercise 3.6. Let $G(n, p_n)$ the Erdos–Rényi graph. In class we saw that as $np_n \to \infty$ then $G(n, p_n)$ contains a triangle with probability tending to one.

Show that when $p \in (0, 1)$ is fixed and T_n denotes the number of triangles in $G(n, p)$, then, *almost surely*,
$$\frac{T_n}{\mathbb{E}[T_n]} \longrightarrow 1.$$

Exercise 3.7 (One-side bound). Let X be a random variable with expectation m and variance σ^2.

1. Show that

$$P(X \geq \lambda) \leq \frac{\sigma^2}{\sigma^2 + (\lambda - m)^2}, \quad \text{for all} \quad \lambda \geq m.$$

   Hint: Apply Markov to $P(f(X) \geq f(\lambda))$ for a well-chosen linear function f.

2. Compare with Chebishev's inequality for $P(|X - m| \geq \lambda - m)$.

Exercise 3.8. The goal of this exercise is to estimate the number of primes that divide an integer smaller than n.

In what follow p denotes a prime number. Let $x \in \mathbb{N}_{>0}$ and set

$$\nu(x) := \text{number of } p \text{ s.t. } p|x.$$

For example $\nu(12) = 2$. We want to prove the following:

Fix an arbitrary slowly increasing sequence $\omega(n) \to \infty$. Then for all but $o(n)$ integers $x \leq n$ we have that

$$|\nu(x) - \log(\log(n))| \leq \omega(n)\sqrt{\log(\log(n))}.$$

In particular

$$\nu(x) \sim \log(\log(n)).$$

We prove this result using the Second moment method.

1. Let $Z(x) := \sum_{p \leq n^{0.1}} \mathbb{1}_{\{p|x\}}$. Note that $|\nu(x) - Z(x)| \leq 10$ for all $x \leq n$.

   We now consider a uniform integer $\boldsymbol{x}$ from 1 to n.

2. Show that $\mathbb{E}[\mathbb{1}_{\{p|\boldsymbol{x}\}}] = \frac{\lfloor n/p \rfloor}{n}$.

3. Using that $\sum_{p \leq t} \frac{1}{p} = \log(\log(t)) + O(1)$, show that

$$\mathbb{E}[Z(\boldsymbol{x})] = \log(\log(n)) + O(1).$$

4. Show that $\mathrm{Var}[Z(\boldsymbol{x})] = \mathbb{E}[Z(\boldsymbol{x})] + O(1)$.

   (Hint: first show that $\mathrm{Cov}(\mathbb{1}_{\{p|\boldsymbol{x}\}}, \mathbb{1}_{\{q|\boldsymbol{x}\}}) \leq \frac{3}{n}$ for every pair of primes $p < q$)

5. Conclude using the Second moment method.

4 Moment method

In the previous section, we have used the first and second moments to get partial information about the asymptotic distribution of a sequence of r.v. (X_n) (estimating $\mathbb{P}(X_n = 0)$, concentration of X_n around its mean). In this section, we will see that, by controlling all moments, we can prove convergence in distribution.

4.1 Convergence in distribution and convergence of moments

For a sequence of r.v. (X_n) and a candidate limit Z, we want to compare the two following properties.

Convergence in distribution $X_n \xrightarrow{d} Z$.

Convergence of moments For any positive integer r, we have $\mathbb{E}[X_n^r] \to \mathbb{E}[Z^r]$.

From convergence in distribution to convergence of moments. Convergence in distribution **does not imply** convergence of moments; consider for example a sequence X_n such that

$$\mathbb{P}(X_n = n) = 1/n = 1 - \mathbb{P}(X_n = 0).$$

Then X_n converges in distribution to 0, while $\mathbb{E}(X_n) = 1$ for all $n \geq 1$.

However, moment convergence holds with additional assumptions.

Proposition 4.1. *Let r be a positive integer and s be a real number with $r < s$. Assume $X_n \xrightarrow{d} Z$ and $\mathbb{E}[|X_n|^s]$ bounded (say by M). Then $\mathbb{E}[X_n^r]$ tends to $\mathbb{E}[Z^r]$.*

In particular, if all (absolute) moments of X_n are bounded, then convergence in distribution implies moment convergence.

Proof. Fix $\varepsilon > 0$. Denoting P_{X_n} the law of X_n, we can write, for any $A > 0$,

$$\mathbb{E}[X_n^r] = \int_{-\infty}^{+\infty} t^r P_{X_n}(dt) = \int_{-\infty}^{-A} t^r P_{X_n}(dt) + \int_{-A}^{A} t^r P_{X_n}(dt) + \int_{A}^{+\infty} t^r P_{X_n}(dt). \tag{11}$$

The first term is bounded as follows

$$\left| \int_{-\infty}^{-A} t^r P_{X_n}(dt) \right| \leq \int_{-\infty}^{-A} \frac{|t|^s}{A^{s-r}} P_{X_n}(dt) \leq \frac{\mathbb{E}[|X_n|^s]}{A^{s-r}} \leq \frac{M}{A^{s-r}}.$$

The third term is bounded by the same quantity. Choosing $A = A_0$ such that $\mathbb{P}(Z = A_0) = 0$ (this will be used later) and $\frac{M}{A_0^{s-r}} \leq \varepsilon$, we have

$$\left| \mathbb{E}[X_n^r] - \int_{-A_0}^{A_0} t^r P_{X_n}(dt) \right| \leq 2\varepsilon.$$

We use the same argument for moments of the limiting variable Z. By Skorohod's representation theorem and Fatou's lemma, we have

$$\mathbb{E}[|Z|^s] \leq \liminf_{n\to\infty} \mathbb{E}[|X_n|^s] \leq M$$

from which we deduce as above:

$$\left|\mathbb{E}[Z^r] - \int_{-A_0}^{A_0} t^r P_Z(dt)\right| \leq 2\varepsilon.$$

We conclude that

$$\left|\mathbb{E}[X_n^r] - \mathbb{E}[Z^r]\right| \leq \left|\int_{-A_0}^{A_0} t^r P_{X_n}(dt) - \int_{-A_0}^{A_0} t^r P_Z(dt)\right| + 4\varepsilon.$$

Since $X_n \xrightarrow{d} Z$, the difference of integrals tends to 0 as n tends to infinity (for fixed A_0), so that

$$\liminf_{n\to\infty} |\mathbb{E}[X_n^r] - \mathbb{E}[Z^r]| \leq 4\varepsilon.$$

Since this holds for any $\varepsilon > 0$, the LHS is zero and $\mathbb{E}[X_n^r]$ tends to $\mathbb{E}[Z^r]$ as claimed. □

From moment convergence to convergence in distribution

Conversely, does convergence of moments imply convergence in distribution? Again, we will need some extra assumption.

Definition 4.2. Let X be a r.v. with finite moments. We say that X is determined by its moments if, for a r.v. Y,

$$\Big(\forall\, r \geq 1,\, \mathbb{E}[Y^r] = \mathbb{E}[X^r]\Big) \Rightarrow (X \overset{d}{=} Y).$$

Theorem 4.3 (Moment method)**.** *Let X_n and X be r.v. such that, for all $r \geq 1$, the convergence $\mathbb{E}(X_n^r) \to \mathbb{E}(X^r)$ holds (in particular, we assume that they have moments of all orders). Assume moreover that X is determined by its moments. Then $X_n \xrightarrow{d} X$.*

The condition "X is determined by its moments" is clearly necessary. If this is not the case, the convergence of moments cannot imply convergence in distribution (this would contradict uniqueness of the limit). It's remarkable that it's also a sufficient condition.

To prove the theorem, we need the notion of tightness.

Definition 4.4. A sequence (X_n) of real-valued random variables is tight if for every $\varepsilon > 0$, there exists a finite interval $I \subset \mathbb{R}$ so that $\mathbb{P}(X_n \in I) \geq 1 - \varepsilon$ for all n. (I depends on ε but not on n.)

Tightness is a kind of compactness for r.v. If X_n converges in distribution, then it is tight. Conversely, a tight sequence always has a subsequence converging in distribution. We also have the following convergence criterion (which we don't prove here).

Proposition 4.5. *Let X_n is a tight sequence and X a r.v. Assume that any subsequence of X_n that converges at all converges to X. Then X_n converges to X in distribution.*

Proof of Theorem 4.3. We first prove the tightness of X_n using Markov's inequality, which yields $\mathbb{P}(|X_n| \geq K) \leq \frac{\mathbb{E}(X_n^2)}{K^2}$. The numerator $\mathbb{E}(X_n^2)$ converges and hence is bounded, so that $\mathbb{P}(|X_n| \geq K)$ tends to 0 as K tends to infinity, uniformly in n, i.e. X_n is tight.

Consider a subsequence $X_{\varphi(n)}$ of X_n that converges in distribution to a random variable Y. We want to prove that $Y \stackrel{d}{=} X$. For any even integer $s > 0$, the sequence $\mathbb{E}[|X_{\varphi(n)}|^s] = \mathbb{E}[X^s_{\varphi(n)}]$ is bounded (since it converges). Therefore, from Proposition 4.1, the moments of $X_{\varphi(n)}$ converge to that of Y. But since $X_{\varphi(n)}$ is a subsequence of X_n, they also converge to the moments of X. By uniqueness of the limit, we conclude that X and Y have the same moment. Since X is determined by its moments, $X \stackrel{d}{=} Y$.

From the above convergence criterion (Proposition 4.5), we conclude that $X_n \xrightarrow{d} X$. □

Uniqueness of measure with given moments.

To apply the above theorem, we need to argue that the limiting distribution is determined by its moments. Luckily there is an easy to check sufficient condition for that.

Proposition 4.6. *Let X with finite moments. If there exists $C > 0$ s.t. $|\mathbb{E}(|X|^r)| \leq C^r r!$ holds for any $r \geq 1$, then X is determined by its moments.*

This applies to most classical distributions: Gaussian, Poisson, geometric, any bounded distribution, ...

Proof. Let Y be a r.v. with the same moment of X; we want to prove that $X \stackrel{d}{=} Y$. For any $r > 0$, we denote $\alpha_r := \mathbb{E}(X^r) = \mathbb{E}(Y^r) < \infty$. We also denote $\beta_k = \mathbb{E}(|X|^k)$ the absolute moments of X (which are a priori not equal to that of Y). As a preparation, we prove that our hypothesis on the usual moments α_k implies an analogue one on the absolute moments.

For even k, absolute and usual moments coincide. For odd k, we use that, for any real number x,

$$|x|^{2k-1} \leq 1 + x^{2k},$$

so that

$$\beta_{2k-1} \leq 1 + \alpha_{2k} \leq 1 + C^{2k}(2k)! \leq (C')^{2k-1}(2k-1)!.$$

The idea is to express the characteristic function of X (and Y) in terms of their moments and to use uniqueness of characteristic functions. Informally

$$e^{itX} \sim \sum_{k=0}^{\infty} (itX)^k/k!$$

so that taking expectation, one has

$$\varphi_X(t) \sim \sum_{k=0}^{\infty} (it)^k \, \alpha_k/k!$$

Rigorously, we have by Taylor theorem the bound:

$$\left| e^{itX} - \sum_{k=0}^{N} (itX)^k/k! \right| \leq \frac{|hx|^{N+1}}{(N+1)!}$$

Taking expectation,

$$\left| \varphi_X(t) - \sum_{k=0}^{N} (it)^k \, \alpha_k/k! \right| \leq \frac{|h|^{N+1}}{(N+1)!} \beta_{N+1} \leq (C'|h|)^{N+1}.$$

Using our bound above for $|h| < \rho := 1/C'$, RHS tends to 0 and

$$\varphi_X(t) = \sum_{k=0}^{\infty} (it)^k \, \alpha_k/k!.$$

A similar formula holds for Y, and therefore $\varphi_X(t) = \varphi_Y(t)$ since they have the same moments. But this only holds on a small interval near 0, namely for $|t| \leq \rho$, while we need that on the whole line to apply the uniqueness theorem for characteristic function!

To get equality on the whole line, we write

$$\left| e^{it_0 X} \left(e^{itX} - \sum_{k=0}^{N} (itX)^k/k! \right) \right| \leq \frac{|hx|^{N+1}}{(N+1)!}$$

so that, for $t < \rho$

$$\varphi_X(t_0 + t) = \sum_{k=0}^{\infty} \mathbb{E}(e^{it_0 X} X^k) \, (it)^k/k!.$$

Denote the expectation $c_{t_0,k}(X)$. These are the successive derivates of φ_X at t_0, so that, for $t_0 < \rho$,

$$c_{t_0,k}(X) = c_{t_0,k}(Y)$$

Hence for $t < \rho$ and $t_0 < \rho$,

$$\varphi_X(t_0 + t) = \varphi_Y(t_0 + t)$$

that is φ_X and φ_Y coincides on $(-2\rho, 2\rho)$. Redoing the same proof shows that they coincide on $(-3\rho, 3\rho)$, and so on.... The characteristic functions φ_X and φ_Y coincide on the whole real line, proving $X \stackrel{d}{=} Y$. □

4.2 Factorial moments and Poisson distribution

We will see a first example of application of the moment method, proving convergence to a Poisson distribution. For this it is often easier to use *factorial moments* than moments.

Definition 4.7. The r-th factorial moment of a r.v. with finite moments is

$$E\ (X)_r = E\big(X(X-1)\cdots(X-r+1)\big).$$

Lemma 4.8. *Let X_n and X be r.v. with finite moments. The following are equivalent:*

- *for any $r \geq 1$, we have $\lim_{n\to\infty} \mathbb{E}[X_n^r] = \mathbb{E}[X^r]$;*
- *for any $r \geq 1$, we have $\lim_{n\to\infty} \mathbb{E}[(X_n)_r] = \mathbb{E}[(X)_r]$.*

Proof. As $(x^r)_{r\geq 0}$ and $((x)_r)_{r\geq 0}$ are both bases of the polynomial ring $\mathbb{Q}[x]$, one has

$$(x)_r = \sum_{k=0}^{r} C_{r,k} x^k, \quad x^r = \sum_{k=0}^{r} D_{r,k}(x)_k,$$

for some constant $C_{r,k}$ and $D_{r,k}$. Thus usual moments are linear combination of factorial moments, and conversely. This implies the lemma. □

Corollary 4.9 (of this Lemma and the moment method)**.** *Let X_n and X be r.v. such that, for all $r \geq 1$, $E\ (X_n)_r \to E\ (X)_r$ (in particular, we assume that they have moments of all orders). Assume moreover that X is determined by its moments. Then $X_n \to_d X$.*

Why factorial moments?

- Poisson distribution has particularly simple factorial moments. Indeed, if X is a Poisson distribution of parameter λ, we recall that its PGF is $P(u) = e^{\lambda(u-1)}$, implying

$$\mathbb{E}[(X)_r] = \frac{d^r}{du^r} P(u)|_{u=1} = \lambda^r.$$

- Factorial moments of sums of indicator variable write nicely

$$\mathbb{E}\left[\left(\sum\nolimits_{\alpha\in A} I_\alpha\right)_r\right] = \sum_{\substack{\alpha_1,\ldots,\alpha_r\in A \\ \textbf{distinct}}} \mathbb{E}\big(I_{\alpha_1}\ldots I_{\alpha_j}\big)$$

Application to triangles in random graph $G(n,p_n)$ for $p_n = c/n$:

As above, we denote by T_n the number of triangles in Erdős-Rényi random graph $G(n,p_n)$. We will, show that, when $\lim_{n\to\infty} np_n = c \in (0,+\infty)$, then T_n converges in distribution to a Poisson random variable.

We start by a purely combinatorial lemma.

Lemma 4.10. *Let I_1, ..., I_r be distinct triangles with vertex sets included in $[n]$. Denote $V(r)$, resp. $E(r)$, the union of their vertex sets, respectively edge sets. Then $|E(r)| \geq |V(r)|$ with equality if and only if the triangles are vertex-disjoint.*

Proof. We proceed by induction on r. For $r = 1$, the statement is trivial.

Let $r \geq 1$ and consider $r + 1$ triangles with vertex sets included in $[n]$. By induction hypothesis we assume $|E(r)| \geq |V(r)|$. We look at the possible intersections of the last triangle I_{r+1} with the previous ones.

- If I_{r+1} has no vertex in common with $I_1,\ldots,I_r$, then it has no edge in common either. We have $|V(r+1)| = |V(r)|+3$ and $|E(r+1)| = |E(r)|+3$, implying $|E(r)| \geq |V(r)|$.
- If I_{r+1} has one vertex in common with $I_1,\ldots,I_r$, then it cannot share an edge with one of these triangles. We have $|V(r+1)| = |V(r)| + 2$ and $|E(r+1)| = |E(r)| + 3$, implying $|E(r)| > |V(r)|$.
- If I_{r+1} has two vertices in common with $I_1,\ldots,I_r$, then it cannot share at most one edge with one of these triangles. We have $|V(r+1)| = |V(r)|+1$ and $|E(r+1)| \geq |E(r)| + 2$, implying $|E(r)| > |V(r)|$.
- If I_{r+1} has all its three vertices in common with $I_1,\ldots,I_r$, it can share up to three edges with one of these triangles. We have $|V(r+1)| = |V(r)|$ and $|E(r+1)| \geq |E(r)|$, implying $|E(r)| \geq |V(r)|$. The equality case occur when each edge of I_{r+1} is already in one of the triangles $I_1,\ldots,I_r$. Since the triangles I_i are assumed to be distinct, they have to be in different triangles. This cannot happen when $I_1,\ldots,I_r$ have disjoint vertex-set.

This concludes the induction step and proves that $|E(r)| \geq |V(r)|$ for all $r \geq 1$. The equality case follows also by induction, using the above discussion. □

Theorem 4.11. *Let $p_n \sim c/n$. Then the number of triangles T_n in a random graph $G(n, p_n)$ tends in distribution towards a Poisson law of parameter $c^3/6$. In particular, $P(T_n = 0) \to \exp(-c^3/6)$ (compare this last statement with what was proved for $np_n \to 0$ and $np_n \to +\infty$).*

Proof. We have already seen that

$$\mathbb{E}[T_n] = \binom{n}{3} p_n^3 \sim \frac{c^3}{6}.$$

Let us consider higher factorial moments

$$\mathbb{E}[(T_n)_r] = \sum_{\substack{I_1,\ldots,I_r \in \binom{[n]}{3} \\ \text{distinct}}} \mathbb{E}[\Delta_{I_1} \ldots \Delta_{I_r}]. \tag{12}$$

By definition of $G(n, p_n)$, we have

$$\mathbb{E}[\Delta_{I_1} \ldots \Delta_{I_r}] = p_n^{|E(r)|}.$$

We denote $G[I_1, \dots, I_r]$ the graph whose vertex and edge sets are the unions of those of $I_1, \dots, I_r$. Then consider $(I_1, \dots, I_r)$ and $(I'_1, \dots, I'_r)$ to be equivalent if $G[I_1, \dots, I_r]$ and $G[I'_1, \dots, I'_r]$ are isomorphic. For fixed r, the number of equivalence classes of r-uples of triangles does not depend on n (for large n). The number of r-uples equivalent to a given $(I_1, \dots, I_r)$ is $O(n^{|V(r)|})$: indeed, to construct an element in this class, we need to choose how to relabel the vertices of $G[I_1, \dots, I_r]$.

Let us split the sum in (12), depending on the equivalence class of r-uples of triangles. The total contribution of the equivalence class of a given $(I_1, \dots, I_r)$ is $O(n^{|V(r)|} p_n^{|E(r)|})$. When $p_n \sim c/n$, this is $O(n^{|V(r)|-|E(r)|})$, which is, from the above Lemma, $o(1)$, unless $(I_1, \dots, I_r)$ are vertex-disjoint. Therefore we have

$$\begin{aligned}\mathbb{E}[(T_n)_r] &= \sum_{\substack{I_1,\dots,I_r \in \binom{[n]}{3} \\ \text{disjoint}}} \mathbb{E}[\Delta_{I_1} \dots \Delta_{I_r}] + o(1) \\ &= \binom{n}{3}\binom{n-3}{3} \dots \binom{n-3r+3}{3} p_n^{3r} + o(1) \sim \frac{c^{3r}}{6^r}.\end{aligned}$$

This is the r-th factorial moment of a Poisson random variable of parameter $c^3/6$. Applying the moment method, we conclude that T_n converges in distribution to a Poisson random variable of parameter $c^3/6$. □

4.3 Cumulants and Normal distribution

We now turn to proving convergence to a Gaussian distribution, using the moment method. In this context, it is easier to use cumulants than moments.

Definition 4.12. Let X be a r.v. with finite moment. Its cumulant are defined by the formal equality

$$\log\left(\mathbb{E}(e^{uX})\right) = \log\left(\sum_{j\geq 0} \frac{\mathbb{E}(X^j)}{j!} u^j\right) = \sum_{j\geq 0} \frac{\kappa_j(X)}{j!} u^j$$

Examples:

- $\kappa_1(X) = \mathbb{E}(X)$, $\kappa_2(X) = \mathrm{Var}(X)$, $\kappa_3(X) = E(X^3) - 3E(X)^2E(X) + 2E(X)^3$.
- If $Z \sim C$ is deterministic, then $\kappa_1(Z) = C$ and $\kappa_r(Z) = 0$ for $r \geq 2$.
- If $Z \sim \mathcal{N}(m, \sigma)$, then it is known that $\mathbb{E}(e^{uZ}) = \exp(mu + \sigma^2 u^2/2)$, from which we deduce

$$\kappa_1(Z) = m,\ \kappa_2(Z) = \sigma^2,\ \kappa_r(Z) = 0 \text{ for } r \geq 3.$$

Some properties:

- If X and Y are independent, we have $\mathbb{E}(e^{u(X+Y)}) = \mathbb{E}(e^{uX})\mathbb{E}(e^{uY})$ and hence
$$\kappa_j(X+Y) = \kappa_j(X) + \kappa_j(Y)$$
In particular, if C is constant, $X+C$ has the same cumulant as X, except the first one (which is shifted by C).

- If $\sum_{j\geq 0} \mathbb{E}(X^j)t^j/j!$ has a non-zero radius of convergence, the equality in the definition is not only formal, but holds for small t.

- In general, we have
$$\kappa_r = E(X^r) + \text{a polynomial in } E(X), E(X^2), \ldots, E(X^{r-1}).$$
We can invert this formula and write
$$\begin{aligned} E(X^r) &= \kappa_r(X) + \text{a polynomial in } E(X), E(X^2), \ldots, E(X^{r-1}) \\ &= \kappa_r(X) + \text{a polynomial in } \kappa_1(X), \kappa_2(X), \ldots, \kappa_{r-1}(X). \end{aligned}$$
For the second equality, we assumed by induction hypothesis that such a formula exists for $E(X), E(X^2), \ldots, E(X^{r-1})$.
These formulas imply that convergence of all moments and convergence of all cumulants are independent (as for factorial moments).

Therefore we have the following criterion for convergence in distribution.

Proposition 4.13 (moment method via cumulants)**.** *Let X and X_n be r.v. with finite moments. Assume X is determined by its moments. If, for all integers $r \geq 1$, $\kappa_r(X_n) \to \kappa_r(X)$. Then X_n tends in distribution towards X.*

Why cumulants?

- Gaussian distributions have easy cumulants.

- Cumulants behave well with respect to independence.

When X_n is a sum of indicator, to compute $\kappa_r(X_n)$, it is useful to consider a multilinear version of cumulants, called joint cumulants (as covariance is a bilinear version of variance).

Definition 4.14 (joint cumulants). Let $X_1, ..., X_r$ be r.v. with finite moments. Their mixed/joint cumulant is defined as
$$\kappa(X_1, \ldots, X_r) = [u_1 \cdots u_r] \log\left(\mathbb{E}(e^{u_1X_1+\cdots+u_rX_r})\right),$$
where $[u_1 \cdots u_r]F$ denotes the coefficients of $u_1 \ldots u_r$ of F, either seen as a formal series in $u_1,\ldots, u_r$ or as an analytic function of $u_1,\ldots, u_r$ around $(0, \ldots, 0)$.

Proposition 4.15. *1. κ_r is a symmetric, multi-linear functional;*

2. $\kappa_r(X) = \kappa(X, ..., X)$ (argument is r times the same variable X) ;

3. $\kappa(X_1, ..., X_r) = 0$ *as soon as the set of variables* $\{X_1, ..., X_r\}$ *can be partitioned in two mutually independent sets.*

Proof. 1. is trivial.

For 2. we write

$$\begin{aligned}\kappa(X, \dots, X) &= [u_1 \cdots u_r \log\left(\mathbb{E}(e^{(u_1+\cdots+u_r)X})\right) \\ &= [u_1 \cdots u_r] \sum_{k=0}^{\infty} \sum_{j \geq 0} \frac{\kappa_j(X)}{j!} (u_1 + \cdots + u_r)^j\end{aligned}$$

We can exchange the infinite sum and coefficient extraction (we're working with formal series). Since $[u_1 \dots u_r](u_1 + \cdots + u_r)^j = \delta_{j,r} j!$, we get

$$\kappa(X, \dots, X) = \kappa_r(X).$$

Let us now prove 3. By symmetry we assume w.l.o.g. that $X_1, ..., X_i$ is independent from $X_{i+1}, ..., X_r$ (for some i between 1 and $r-1$). Then

$$\log\left(\mathbb{E}(e^{i(t_1X_1+\cdots+t_rX_r)})\right) = \log\left(\mathbb{E}(e^{i(t_1X_1+\cdots+t_iX_i)})\right) + \log\left(\mathbb{E}(e^{i(t_{i+1}X_{i+1}+\cdots+t_rX_r)})\right)$$

The coefficient of $t_1 \cdots t_r$ is zero, since the first part does not depend on $t_{i+1}, \dots, t_r$ and the second not on $t_1, \dots, t_i$. □

We also note that there are formulae relating joint cumulants and joint moments, for example

$$\begin{aligned}\kappa(X) = \mathbb{E}(X), \quad \kappa(X, Y) &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathrm{Cov}(X, Y), \\ \kappa(X, Y, Z) &= \mathbb{E}(XYZ) - \mathbb{E}(XY)\mathbb{E}(Z) - \mathbb{E}(XZ)\mathbb{E}(Y) - \mathbb{E}(YZ)\mathbb{E}(X) + 2\mathbb{E}(X)\mathbb{E}(Y)\mathbb{E}(Z); \\ \mathbb{E}(X) = \kappa(X), \quad \mathbb{E}(XY) &= \kappa(X, Y) + \kappa(X)\kappa(Y), \\ \mathbb{E}(XYZ) &= \kappa(X, Y, Z) + \kappa(X, Y)\kappa(Z) + \kappa(X, Z)\kappa(Y) + \kappa(Y, Z)\kappa(X) + \kappa(X)\kappa(Y)\kappa(Z).\end{aligned}$$

The general form uses the combinatorics of set-partitions. The detail will not be useful here, but we shall use the following consequence.

Lemma 4.16. *For any $r > 0$, there exists a constant $B_r > 0$ such that, for any r.v. $X_1, \dots, X_r$ bounded by 1 (e.g. indicators), we have*

$$|\kappa(X_1, \dots, X_r)| \leq B_r.$$

Proof. Joint cumulants can be written in terms of joint moments with coefficients which do not depend of the variables $X_1, \dots, X_r$. But joint moments of variables bounded by 1 are bounded by 1 themselves. This ends the proof. □

Application to number of triangles in $G(n, p)$ As above, let T_n be the number of triangles in $G(n, p)$; (for simplicity, we take $p_n = p$ constant). We set

$$\hat{T_n} = \frac{T_n - \mathbb{E}(T_n)}{\sqrt{\mathrm{Var}(T_n)}},$$

and recall that $\mathrm{Var}(T_n) = \Theta(n^4)$.

We analyse the cumulants of T_n.

$$\kappa_r(T_n) = \kappa_r\left(\sum_{I\in\binom{[n]}{3}} \Delta_I\right) = \sum_{I_1,\dots,I_r\in\binom{[n]}{3}} \kappa\big(\Delta_{I_1},\dots,\Delta_{I_r}\big) \tag{13}$$

As for factorial moments, it is clear that the summand $\kappa\big(\Delta_{I_1},\dots,\Delta_{I_r}\big)$ only depends of the isomorphism class of the graph $G[I_1,\dots,I_r]$.

Call (C) the property "the graph $G[I_1,\dots,I_r]$ is *edge-connected* in the sense, that it cannot be split into two collections of triangles so that no triangles of the first collection share an edge with a triangle of the second collection". If (C) is not satisfied, then the corresponding random variables $\Delta_{I_1},\dots,\Delta_{I_r}$ can be split into two collections of mutually independent random variables and the cumulant $\kappa\big(\Delta_{I_1},\dots,\Delta_{I_r}\big)$ is 0. Therefore the above sum can be restricted to r-uples $(I_1,\dots,I_r)$ satisfying (C).

Claim: under condition (C), the graph $G[I_1,\dots,I_r]$ has at most $r+2$ vertices.

Proof: it is easy to see that, assuming condition (C), there exists a permutation $J_1,\dots,J_r$ of $I_1,\dots,I_r$ such that for each $s\le r$, the family J_{s+1} shares an edge with $(J_1,\dots,J_s)$. Indeed, we construct $J_1,\dots,J_s$ greedily: choose J_1 as you want, and if $(J_1,\dots,J_s)$ is constructed, the property (C) ensures you that one of the not yet selected I_t's share an edge with $(J_1,\dots,J_s)$. Then an immediate induction proves that, for any $s\le r$, the graph $G[J_1,\dots,J_s]$ has at most $s+2$ vertices, implying in particular the claim.

We deduce that the claim that, under condition (C), the number of r-tuples in the same equivalence class than $(I_1,\dots,I_r)$ is $O(n^{r+2})$. Since, $\kappa\big(\Delta_{I_1},\dots,\Delta_{I_r}\big)$ is bounded by a universal constant B_r (see above lemma), the total contribution of that equivalence class to (13) is $O(n^{r+2})$. Since the number of equivalence classes does not depend on n (for large n), we have

$$\kappa_r(T_n) = O(n^{r+2}).$$

Recall that shifting a variable by C does not change its cumulants (except the first one), and multiplying by λ multiplies the r-th cumulant by λ^r. This implies that, for $r\ge 2$,

$$\kappa_r(\hat{T_n}) = \frac{\kappa_r(T_n)}{\sqrt{\mathrm{Var}(T_n)}^r} = O(n^{-r+2}).$$

In particular $\kappa_r(\hat{T_n})$ is 1 for $r=2$ and tends to 0 for $r>2$, implying that $\hat{T_n}$ converges in distribution to a standard Gaussian random variable. This proves Theorem 1.5, item iii) from the introduction in the case $p_n = p$.

Comment: In fact, T_n converges after renormalization to a standard Gaussian distribution as soon as $np_n\to\infty$ and $n^2(1-p_n)\to\infty$ (recall that the above proof only considers the case where $p_n = p$ is constant). This can also

be proved using cumulants, but the combinatorial analysis of cumulants is then more delicate. In particular, the bound $|\kappa_r(\Delta_{H_1}, \dots, \Delta_{H_r})| \leq B_r$ is in general too coarse.

Also, similar results can be proved for the number of copies of other subgraphs than triangles. See Janson, Łuczak, Ruçinski, *Random Graphs*, Wiley Interscience, 2000.

4.4 Exercices

Exercise 4.1. We consider the following situation: we have two uniformly shuffled decks with $4n$ cards (n for each of the 4 possible suits). We turn the card of the two decks one by one (one from each deck at each time) and we count the number of times you get the same value, but possibly with different suits. We denote this number as X_n.

1. Note that X_n is distributed as the number of i such that $\sigma(i) \equiv i \mod n$ in a uniform permutation σ of size $4n$.
2. Write X_n as a sum of indicator functions.
3. We recall that factorial moments of sums of indicator variables write nicely:
$$\mathbb{E}\left[\left(\sum\nolimits_{\alpha\in A} I_\alpha\right)_r\right] = \sum_{\substack{\alpha_1,\dots,\alpha_r\in A\\ \textbf{distinct}}} \mathbb{E}\big(I_{\alpha_1}\dots I_{\alpha_r}\big).$$
Using this result, write the r-th the factorial moment of X_n as:
$$\sum_{\substack{i_1,\dots,i_r\in[4n]\\ \textbf{distinct}}} \mathbb{P}(\sigma(i_1)\equiv i_1 \mod n, ..., \sigma(i_r)\equiv i_r \mod n). \qquad (14)$$
4. Show that
$$\mathbb{P}(\sigma(i_1)\equiv i_1 \mod n, ..., \sigma(i_r)\equiv i_r \mod n) \leq \frac{4^r(4n-r)!}{(4n)!}.$$
5. Show that the number of terms in Equation (14) where some i_j's are congruent to each other is $O(n^{r-1})$.
6. Conclude that X_n converges to a Poisson random variable of parameter 4.

Exercise 4.2. Let X and Y be two real-valued random variables with densities:
$$f_X(x) = \frac{e^{-\log(x)^2/2}}{x\sqrt{2\pi}}\mathbf{1}_{x>0},$$
$$f_Y(x) = \frac{e^{-\log(x)^2/2}}{x\sqrt{2\pi}}\big(1+\sin(2\pi\log(x))\big)\mathbf{1}_{x>0}.$$

(Why does this define random variables?)

Show that X and Y have the same moments but distinct distributions.

Hint: compute the r-th moments of X and Y as an integral and do the change of variables $x = \log(t + r)$.

Exercise 4.3. In this exercise we reprove the central limit theorem for the number X_n of descents in uniform random permutation using cumulants. You can assume that (this was proved during the course)

$$\mathrm{Var}(X_n) = \frac{n}{12} + O(1).$$

1. Write X_n as a sum of indicator functions.
2. Express the r-th cumulant of X_n as a sum of simpler cumulants.
3. Note that most of the simpler cumulants are equal to 0. Which ones? Why?
4. Estimate the number of simpler cumulants that are different from zero.
5. Prove a central limit theorem for X_n.

5 Azuma inequality

In this chapter we discuss again concentration inequalities, i.e. upper bounds for $\mathbb{P}\big[|X_n - \mathbb{E}(X_n)| \geq \varepsilon \mathbb{E}(X_n)\big]$. Assume that X_n is asymptotically normal with expectation and variance of order n (e.g. the already discussed example of the number of descents in uniform random permutations).

Then by Chebyshev's inequality, we get, for fixed ε,

$$\mathbb{P}\big[|X_n - \mathbb{E}(X_n)| \geq \varepsilon \mathbb{E}(X_n)\big] \leq \frac{1}{\varepsilon^2} \frac{\mathrm{Var}(X_n)}{\mathbb{E}(X_n)^2} = O(n^{-1}).$$

If we have convergence of the moments of $(X_n - \mathbb{E}X_n)/\sqrt{\mathrm{Var}(X_n)}$ to those of the Gaussian distribution, then we can improve this bound by applying Markov's inequality to $(X_n - \mathbb{E}(X_n))^r / \mathrm{Var}(X_n)^{r/2}$. For even r, we have

$$\begin{aligned}\mathbb{P}\big[|X_n - \mathbb{E}(X_n)| \geq \varepsilon \mathbb{E}(X_n)\big] &= \mathbb{P}\left[\frac{(X_n - \mathbb{E}(X_n))^r}{\mathrm{Var}(X_n)^{r/2}} \geq \frac{\varepsilon^r \mathbb{E}(X_n)^r}{\mathrm{Var}(X_n)^{r/2}}\right] \\ &\leq \mathbb{E}\left[\frac{(X_n - \mathbb{E}(X_n))^r}{\mathrm{Var}(X_n)^{r/2}}\right] \frac{\mathrm{Var}(X_n)^{r/2}}{\varepsilon^r \mathbb{E}(X_n)^r} = O(n^{-r/2}).\end{aligned}$$

In the last estimate, we have used that $\mathbb{E}\left[\frac{(X_n - \mathbb{E}(X_n))^r}{\mathrm{Var}(X_n)^{r/2}}\right] = O(1)$, (because it converges). We have proved that the deviation probability decays faster than any polynomial in n.

In case of descents in uniform random permutations, we have seen that it decays exponentially fast in n, using Chernoff's bound. This relies however on the computation of the PGF. In this section, we will see a simple way to find exponential bounds, without computing the PGF.

The theorem

Definition 5.1. A family $\{D_i, i \in I\}$ is called a *multiplicative system* if for any non-empty subset $J \subseteq I$, we have $\mathbb{E}\left(\prod_{j \in J} D_j\right) = 0$.

We note for later use that, for a multiplicative system $\{D_i, i \in I\}$ and families $\{a_i, i \in I\}$ and $\{b_i, i \in I\}$ of scalars, one has

$$\mathbb{E}E\left(\prod_{i \in I}(a_i + b_i D_i)\right) = \prod_{i \in I} a_i = \prod_{i \in I} \mathbb{E}E(a_i + b_i D_i),$$

as if the D_i were independent, but only for linear functionals.

Theorem 5.2 (Azuma's inequality)**.** *Let $\{D_i, i \in I\}$ be a multiplicative system of* bounded *r.v., i.e. for any $i \in I$, there exists M_i in $\mathbb{R}$ such that $|D_i| \leq M_i$ almost surely. Then, for any $\lambda > 0$,*

$$P\left(\left|\sum_{i \in I} D_i\right| > \lambda\right) \leq 2 \exp\left(-\frac{\lambda^2}{2\sum_i M_i^2}\right).$$

Note: there is no asymptotics in the statement. When we apply it to a sequence of multiplicative families, we can have different λ and M_i for each of these families, i.e. λ and M_i can depend on n.

Proof. Let $X = \sum_{i\in I} D_i$. We will use Chernoff's inequality. For this we need to control

$$\mathbb{E}\big[\exp(uX)\big] = \mathbb{E}\left[\prod_{i\in I} \exp(uD_i)\right].$$

The D_i are not assumed to be independent, so that we cannot exchange the product and the expectation. We need to control $\exp(uD_i)$ by a linear functional.

For any fixed t, x with $|x| \le 1$, using the convexity of e^{tx}, we have

$$e^{tx} \le \cosh t + x \sinh t.$$

Setting $x = D_i/M_i$ and $t = M_i\, u$, we get

$$\exp(uD_i) \le \cosh(M_i\, u) + \frac{D_i}{M_i} \sinh(M_i\, u).$$

Taking the product over i, the expectation, and using properties of multiplicative system, we have

$$\mathbb{E}\left[\prod_{i\in I} \exp(uD_i)\right] \le \mathbb{E}\left[\prod_{i\in I} (\cosh(M_i\, u) + \frac{D_i}{M_i} \sinh(M_i\, u))\right] = \prod_{i\in I} \cosh(M_i\, u)$$

By Chernoff's bound, for $u > 0$,

$$\mathbb{P}\left(\sum_{i\in I} D_i > \lambda\right) \le \exp(-u\lambda) \prod_{i\in I} \cosh(M_i\, u).$$

To find the bound of the theorem, we use that $\cosh(x) \le \exp(x^2/2)$ for any $x \in \mathbb{R}$ (which can be proved by comparing term by term their power series expansion) and optimize over u:

$$\mathbb{P}\left(\sum_{i\in I} D_i > \lambda\right) \le \inf_{u>0} \exp\Big(-u\lambda + \tfrac{1}{2}\sum_{i\in I} M_i^2 u^2\Big) = \exp\left(-\frac{\lambda^2}{2\sum_{i\in I} M_i^2}\right).$$

By symmetry ($-D_i$ is also a bounded multiplicative family), the same bounds holds □

How to construct multiplicative systems?

Imagine your random space is built by taking independent random variables $(A_1, ..., A_N)$. *i.e.,* $\Omega = \Omega_1 \times \Omega_2 \cdots \times \Omega_N$

Examples: uniform random words, where each A_i is a letter, random graphs $G(n,p)$, where each A_i is the indicator of the presence of an edge, uniform random permutations constructing by standardizing independent uniform variables in $[0,1]$, where each A_i is one of these random variable.

We want to study some statistics X on such a probability space.

For $0 \leq i \leq n$, consider the following random variable:

$$\begin{aligned} F_i(A_1, ..., A_n) &= \mathbb{E}(X|A_1, ..., A_i) \\ &= \mathbb{E}(X(A_1, ..., A_i, A'_{i+1}, ..., A'_n)). \end{aligned} \tag{15}$$

In the second line, A_1,...,A_i are fixed and the expectation is taken with respect to $A'_{i+1}, \ldots, A'_n$ which are independent from each other and have the distribution of $A_{i+1}, \ldots, A_n$. Note that F_i depends in fact only of $A_1, ...A_i$. For example, $F_0 = \mathbb{E}(X(A'_1, ..., A'_n))$ is a real number, while $F_n = X(A_1, ..., A_n)$ is the original random variable X.

For i in $\{1, \ldots, n\}$, we set $D_i = F_i - F_{i-1}$. Note that $\sum_{i=1}^n D_i = F_n - F_0 = X - \mathbb{E}(X)$.

Lemma 5.3. *The family $(D_i)_{1 \leq i \leq n}$ is a multiplicative system.*

Proof. In fact, we shall prove a more general statement: if G depends only on $A_1, ..., A_{i-1}$, then

$$\mathbb{E}\big(G(A_1, ..., A_{i-1})D_i\big) = 0. \tag{16}$$

Proof of (16): we have

$$\begin{aligned} \mathbb{E}\big(G(A_1, \ldots, A_{i-1})D_i\big) &= \mathbb{E}\Big[\mathbb{E}\big(G(A_1, \ldots, A_{i-1})D_i|A_1, \ldots, A_{i-1}\big)\Big] \\ &= \mathbb{E}\Big[G(A_1, \ldots, A_{i-1})\big(\mathbb{E}(F_i|A_1, \ldots, A_{i-1}) - F_{i-1}\big)\Big] = 0 \end{aligned}$$

Why is (16) *stronger than the multiplicative system property?* Let $J \subset \{1, \ldots, n\}$ be non-empty, we need to prove that $\mathbb{E}\big(\prod_{i \in J} D_i\big) = 0$. Let $i_0 = \max(J)$. We have

$$\mathbb{E}\left[\prod_{i \in J} D_i\right] = \mathbb{E}\left[\underbrace{\left(\prod_{i \in J \setminus \{i_0\}} D_i\right)}_{\text{this is a } G(A_1, ..., A_{i_0 - 1})} D_{i_0}\right] = 0 \qquad \square$$

Comment: (16), with the fact that F_i only depends on $A_1, \ldots, A_i$, says that F_i is a martingale with respect to the obvious filtration associated to the (A_i); and indeed (15) is a standard way of constructing martingale. The rest of the proof shows that differences of martingale are multiplicative systems.

Azuma's inequality assumes the D_i to be bounded a.s. How can we check this assumption when D_i is constructed as above? Recall that

$$D_i = F_i - F_{i-1} = \mathbb{E}\big[X(A_1, ..., A_i, A'_{i+1}, ..., A'_n) - X(A_1, ..., A'_i, A'_{i+1}, ..., A'_n)\big].$$

In particular if X changes at most by M_i when the value of its i-th argument changes, then $D_i \leq M_i$ a.s. and we can apply Azuma's inequality. Such a property is usually straight-forward to check.

Application to the longest common substring problem. As in the introduction, let Ω be a finite alphabet with k letters (k fixed) and w and w' be two independent random words of length n with letters in Ω. We are interested in the length L_n of the longest common substring between w and w'. Recall from the introduction that $\mathbb{E}(L_n) \sim \gamma_k n$.

We set, for i between 0 and n (or 0 and $n-1$)

$$F_{2i} = \mathbb{E}\big[L_n | w_1, w'_1, \ldots, w_i, w'_i\big];$$
$$F_{2i+1} = \mathbb{E}\big[L_n | w_1, w'_1, \ldots, w_i, w'_i, w_{i+1}\big].$$

Furthermore, let $D_i = F_i - F_{i-1}$. As explained above $\sum_{i=1}^n D_i = L_n - \mathbb{E}(L_n)$ Since all the letters of w and w' are independent from each other, the family $(D_i)_{1 \leq i \leq 2n}$ is multiplicative.

Moreover, changing one letter in one of the word can at most change L_n by 1, so that $|D_i| \leq 1$ almost surely. From Azuma's inequality, we get, that for any $t_n > 0$,

$$\mathbb{P}\big[|L_n - \mathbb{E}(L_n)| \geq t_n\big] \leq 2\, \exp(-t_n^2/4n).$$

This proves the second part of Theorem 1.3 from the introduction.

Azuma's inequality has a large range of applications in (and outside) combinatorics and is usually straight-forward to apply. Further applications can be found in the exercises.

5.1 Exercises

Exercise 5.1. Let T_n be the number of triangles in the random Erdős-Rényi graph $G(n,p)$ (p might depend on n). Show that, for any $t = t_n$, we have

$$\mathbb{P}\big[|Z_n^p - \mathbb{E}(Z_n^p)| \geq t_n\big] \leq \exp\big(-\tfrac{t_n^2}{n^4}\big).$$

Exercise 5.2. Let Z_n^p be the chromatic chromatic number of the random Erdős-Rényi graph $G(n,p)$ (p might depend on n). Show that, for any $t = t_n$, we have

$$\mathbb{P}\big[|Z_n^p - \mathbb{E}(Z_n^p)| \geq t_n\big] \leq \exp\big(-\tfrac{t_n^2}{2n}\big).$$

A Scratch course/reminder of complex analysis

Notation: $U \subseteq \mathbb{C}$ an open set, $D(z_0, r) = \{z : |z - z_0| < r\}$ an open disk.

Definition A.1 (holomorphic). We say that $f : U \to \mathbb{C}$ is holomorphic on Uif one/both of the following equivalent conditions hold:

1. for every $z_0 \in U$, the limit
$$\lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}$$
exists.

2. for every $z_0 \in U$ and every $r > 0$ s.t. $D(z_0, r) \subseteq U$ there exists a sequence $(a_n)_{n \geq 0}$ of complex numbers such that $f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$ for all $z \in D(z_0, r)$ (in particular, the RHS should converge for such z).

Note: equivalence is highly nontrivial and very surprising (item 2. implies being infinitely many times differentiable and thus seems much stronger) !

Standard functions (polynomials, rational functions, exponential, linear combinations, products, quotients, compositions of those) are holomorphic on their domain of definition. The logarithm and the noninteger power functions ($z \mapsto z^\alpha$, with $\alpha \notin \mathbb{Z}$) can be extended to holomorphic functions to the *split plane* $\mathbb{C} \setminus \mathbb{R}_{\leq 0}$ (or to any set U obtained by removing from $\mathbb{C}$ a closed half-line ending at 0).

Path integrals

Definition A.2 (path). A path γ is a continuous piecewise-C^1 function from a real bounded closed interval $[a, b]$ to $\mathbb{C}$.

Definition A.3 (path integrals). We consider a continuous function $f : U \to \mathbb{C}$ and a path $\gamma : [a, b] \to U$. Then we define
$$\int_\gamma f(z)dz := \int_a^b f(\gamma(t))\gamma'(t)dt.$$

Proposition A.4. *One has* $\left|\int_\gamma f(z)dz\right| \leq \sup_{t \in [a;b]} |f(\gamma(t))| \cdot L(\gamma)$, *where, by definition* $L(\gamma) := \int_a^b |\gamma'(t)|dt$ *is the* length *of the path* γ.

The residue theorem. We now present the residue theorem, which evaluates path integrals of quotients of holomorphic functions on a closed path (*i.e.,* path with $\gamma(b) = \gamma(a)$).

Definition A.5.
- A meromorphic function on U is a quotient $h(z) = f(z)/g(z)$ of holomorphic functions on U (h might not be defined on the whole set U).

- A point $a \in U$ is a *pole* of h if $\lim_{z\to a} |h(z)| \to +\infty$.

Note: a necessary condition for a to be a pole is that $g(a) = 0$, but it is not sufficient (if $f(a) = 0$ as well, the limit $\lim_{z\to a} |h(z)|$ might be finite). By the isolated zero principle, the set of zeroes of g, and hence the set of poles of h has no limit points in U.

Proposition A.6. *Let $h(z)$ be a meromorphic function on U and $a \in U$ be a pole of h. Then there exists $r > 0$, an integer $m > 0$ and coefficients $(b_n)_{n\geq -m}$ such that $D(a, r) \subseteq U$ and*

$$h(z) = \sum_{n=-m}^{+\infty} b_n (z-a)^n.$$

Definition A.7 (order and residue). The minimal m such that this expansion exists is called the *order* of the pole and the coefficient b_{-1} is the **residue** of h in a, denoted $\mathrm{Res}(h; a)$.

Note: if the limit $\lim_{z\to a}(z-a)h(z)$ exists and is non-zero (and finite), then $m = 1$ (the pole is called **simple**) and $\mathrm{Res}(h; a) = \lim_{z\to a}(z-a)h(z)$.

We now state the residue theorem: to simplify consider a closed path γ, injective on $[a, b)$. It separates the plane into two regions: the interior V and the exterior $\mathbb{C} \setminus \overline{V}$ (with this notation, $\gamma([a, b]) = \partial V$). We also assume γ to be oriented counterclockwise. Such paths are called (positive) *contour*. Contour integral are sometimes denoted by $\oint$.

Theorem A.8 (Residue theorem)**.** *Let h be a meromorphic function on U and γ a contour in U, so that no poles of h are on the path γ. Then*

$$\oint_\gamma h(z)dz = 2\pi\, i \sum_{\substack{a\in V \\ a \text{ pole of } h}} \mathrm{Res}(h; a).$$

Corollary A.9 (Cauchy formula for derivatives)**.** *Let $f(z)$ be a* **holomorphic** *function on U. We fix z_0 in U and a contour γ having z_0 in its interior. Then*

$$\frac{f^{(k)}(z_0)}{k!} = \frac{1}{2\pi\, i} \oint_\gamma \frac{f(z)}{(z-z_0)^{k+1}} dz.$$

Controlling derivatives A consequence is that, if we have some bounds on a sequence of holomorphic functions f_n on U, we can control its derivatives.

Proposition A.10. *Let U be an open set containing 0 and $f_n : U \to \mathbb{C}$ be a sequence of holomorphic functions. Assume there exist real numbers $C_n, r > 0$ and an integer $d \geq 0$ such that, for $|z| < r$, we have $z \in U$ and*

$$|f_n(z)| \leq C_n z^d.$$

Then for any $r' < r$, we have

$$f_n^{(k)}(0) \leq C_n (r')^{d-k}.$$

In particular, if $d > k$, then $f_n^{(k)}(0) = 0$.

Sketch of proof. Easy consequence of Cauchy formula for derivatives (using the circle of radius r' as contour path) and of the standard estimate for path integrals.

For the last statement, we simply make r' tends to 0. □

Application 1: If $f_n(z) = O(z^2)$ for $z \to 0$ uniformly in n, then $f_n'(0) = 0$ and $f_n''(z) = O(1)$. We can use this to simplify the computation p. 9.

Application 2: If $f_n(z) = o(1)$ (resp. $O(1)$) for $n \to \infty$ uniformly in z in a neighbourhood of 0, then $f_n^{(k)}(z_0) = o(1)$ (resp. $O(1)$) uniformly on z_0 in a (smaller) neighbourhood of 0 (applying the above proposition to $f_n(z + z_0)$). This is used (only for $z_0 = 0$) in the proof of Theorem 2.3.

B Sources

Various sources have been used for the preparation of this lecture, in particular

- Chapter IX of the wonderful book of Flajolet and Sedgewick *Analytic Combinatorics* [FS09] for Section 2;
- Chapters 2, 4 and 7 of the marvellous volume of Alon and Spencer *The Probabilistic Method* [AS15] for Sections 3 and 5;
- Chapter 6 of the beautiful opus of Janson, Łuczak and Rucinski *Random Graphs* [JŁR00] for Section 4 (except the proof of the moment method which is taken from Billingsley's classical and always enjoyable textbook [Bil12, Chapter 20]).

References

[AS15] N. Alon, J. Spencer, *The probabilistic method*, 4th Edn, Wiley series in discrete mathematics and optimization, Wiley, 2015.

[Bil12] P. Billingsley, *Probability and Measure*, Anniversary Edn, Wiley Intersciences, 2012.

[FS09] Ph. Flajolet, R. Sedgewick. *Analytic combinatorics*. Cambridge University Press, 2009.

[JŁR00] S. Janson, T. Łuczak, A. Rucinski, *Random graphs*. Wiley Intersciences, 2000.